

Duální řízení Markovských řetězců

Dual Control of Markov Chain

Zadání diplomové práce

Student:

Bc. Dušan Polek

Studijní program:

N2647 Informační a komunikační technologie

Studijní obor:

1103T031 Výpočetní matematika

Téma:

Duální řízení Markovských řetězců

Dual control of Markov chains

Zásady pro vypracování:

Cílem práce je návrh optimální rozhodovací strategie pro jednoduchý diskrétní náhodný proces s pamětí a řídicími vstupy, jehož parametry nejsou přesně známy.

Příkladem takového procesu mohou být opakované hody nesymetrickou mincí, přičemž výsledek hodu (výstup) bude záviset na tom, jak byla mince položena před hodem (vstup) a jak dopadl poslední hod (paměť). Parametry mince (tj. pravděpodobnosti hodu pany/orla pro jednotlivé kombinace vstupu a posledního výsledku) přitom nejsou známy, nebo jsou známy jen částečně. Naším cílem bude najít optimální rozhodovací strategii (posloupnost funkcí přiřazujících v každém čase optimální vstup k posloupnosti předchozích vstupů a výstupů) tak, abychom v určitém časovém horizontu získali např. maximální počet hlav.

Optimální strategii pro takovýto případ lze nalézt na základě tzv. bayesovské teorie. Takto získaná strategie má tu zajímavou vlastnost, že v průběhu řízení dochází zároveň k aktivnímu učení (získávání a využívání nových informací o systému). Může se tak stát, že řídicí systém v určitých časech navrhne rozhodnutí, které je z krátkodobého hlediska chybné (nižší pravděpodobnost hlavy v následujícím okamžiku), zato ale povede k získání cenné informace o systému. To umožní zlepšit řízení v následujících krocích a celkově tak povede k vyššímu celkovému zisku.

Takovýto způsob řízení se nazývá duální (kombinuje samotné řízení s procesem učení). Rozhodovací strategie se pak hledá pomocí dynamického programování, což (bez použití aproximací) vede k výpočetně náročným úlohám i pro velmi jednoduché systémy.

Pokyny k vypracování:

1. Seznamte se základy bayesovské teorie pro dynamické rozhodování za neurčitosti.
2. Sestavte a implementujte algoritmus pro nalezení optimální rozhodovací strategie pro výše popsany systém.
3. Pokuste se nalézt apriorní distribuci (reprezentují počáteční znalost systému) takovou, že dojde situaci, kdy optimální strategie navrhne krátkodobě nevýhodná rozhodnutí.

Seznam doporučené odborné literatury:

Nagy I., Pavelková L., Suzdaleva E., Homolová J., Kárný M. BAYESIAN DECISION MAKING Theory and Examples, UTIA AVCR, 2005

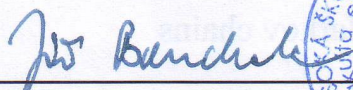
Další literatura po dohodě se školitelem.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Jan Kracík, Ph.D.**

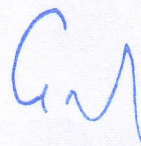
Datum zadání: 01.09.2013

Datum odevzdání: 07.05.2014



doc. RNDr. Jiří Bouchala, Ph.D.
vedoucí katedry





prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 4. května 2014

.....

Rád bych na tomto místě poděkoval vedoucímu mé diplomové práce Ing. Janu Kracíkovi, Ph.D. za čas, který věnoval k poskytování mi spousty cenných rad a informací.

Abstrakt

Práce se zabývá návrhem optimální rozhodovací strategie jednoduchého Markovského rozhodovacího procesu, jehož parametry nejsou přesně známy. Optimální strategii hledáme na základě bayesovské teorie. Optimalizační proces se pak skládá ze dvou vzájemně propojených částí, návrhu strategie a aktivního učení (získávání a využívání nových informací o systému). Takovýto způsob řízení se nazývá duální řízení.

Klíčová slova: Markovský rozhodovací proces, bayesovská statistika, duální řízení

Abstract

This thesis is engaged in design of optimal decision strategy of simple Markov decision process, whose internals aren't known exactly. The optimal strategy is designed using Bayesian statistics. Thus, the optimization procedure is taken as consisting of two interconnected parts, strategy design and active learning (acquiring and using new informations about the system). This control method is referred to as a dual control.

Keywords: Markov decision process, Bayesian statistics, dual control

Seznam použitých zkratk a symbolů

θ	– parametr
$f(\theta)$	– apriorní hustota parametru θ
$f(\theta x)$	– posteriorní hustota parametru θ
$P(x y)$	– podmíněná pravděpodobnost x za podmínky y
t	– diskrétní časový okamžik
x_t	– hodnota veličiny v čase t
$t : T$	– časová posloupnost $(t, t + 1, \dots, T)$
$x_{t:T}$	– posloupnost veličin $(x_t, x_{t+1}, \dots, x_T)$
$:=$	– definiční rovnost

Obsah

1	Úvod	5
2	Markovské řetězce	7
2.1	Markovský řetězec jako náhodný proces	7
2.2	Přechodová matice	7
2.3	Absorpční řetězec	8
2.4	Příklady Markovského řetězce	8
2.5	Markovský rozhodovací proces	9
3	Vybraná rozdělení pravděpodobnosti	12
3.1	Alternativní rozdělení	12
3.2	Binomické rozdělení	12
3.3	Rovnoměrné rozdělení	13
3.4	Beta rozdělení	13
3.5	Další rozdělení pravděpodobnosti	15
4	Bayesovské metody	16
4.1	Bayesův vzorec	16
4.2	Volba apriorního rozdělení	17
4.3	Statistické rozhodovací funkce	22
4.4	Dynamický Bayesovský rozhodovací proces	24
5	Formulace problému	29
5.1	Model	29
5.2	Rozhodovací úloha	30
5.3	Aplikace vlastností modelu na rozhodovací úlohu	31
5.4	Konkrétní příklad	36
5.5	Optimální strategie navrhuující krátkodobě nevýhodná rozhodnutí	43
5.6	Další příklady	49
6	Závěr	50
7	Reference	51

Seznam tabulek

1	Další rozdělení pravděpodobnosti	15
2	Konjugované apriorní hustoty některých modelů	21
3	Průměrné ztráty optimální strategie (PZO) a suboptimální strategie (PZS) pro různé skutečné parametry mince a zvolené parametry apriorního beta rozdělení.	49

Seznam obrázků

1	Markovský řetězec představující hod pravidelnou symetrickou mincí. . .	8
2	Markovský řetězec představující jednoduchou hru o body.	9
3	Markovský rozhodovací proces znázorňující hod nesymetrickou mincí. . .	11
4	Průběh funkce hustoty pravděpodobnosti beta rozdělení pro různé parametry.	14
5	Náš uvažovaný model jako Markovský rozhodovací proces, kde pravděpodobnosti přechodu závisí na zvolené akci.	30
6	Ztráty optimální strategie v jednotlivých pokusech pro 1000 simulací. . . .	41
7	Četnosti ztrát v 1000 simulovaných pokusech optimální strategie.	42
8	Ztráty suboptimální strategie v jednotlivých pokusech pro 1000 simulací.	44
9	Četnosti ztrát v 1000 simulovaných pokusech suboptimální strategie. . . .	45
10	Srovnání četností ztrát v 1000 simulovaných pokusech optimální a suboptimální strategie.	46

Seznam výpisů zdrojového kódu

1	Vstupní parametry algoritmu pro hledání optimální strategie	37
2	Funkce pro převod indexu vektoru v pro čas t na trajektorii	37
3	Funce učení počítající parametry aposteriorního rozdělení	37
4	Algoritmus hledání optimální strategie	38
5	Simulování výsledků hodů mincí	39

1 Úvod

Tato práce se zabývá návrhem optimální rozhodovací strategie pro jednoduchý diskrétní náhodný rozhodovací proces, jehož parametry přesně neznáme. Příkladem takového procesu může být opakovaný hod mincí. Obvykle se setkáváme s tvrzením, že pravděpodobnost hodu panny či orla je stejná, ve skutečnosti je však hod mincí deterministický proces určený mnoha okolnostmi průběhu daného hodu. Těmito okolnostmi může být např. poloha mince před hodem, síla, kterou minci vrhneme, velikost mince a mnoho dalších.

My však budeme uvažovat jednodušší příklad náhodného procesu. Budeme se zabývat opakovanými hody nesymetrickou mincí, jejíž výsledek hodu (padne panna nebo orel) bude záviset na tom, v jaké poloze byla mince před hodem (vstup) a jak dopadl poslední hod (paměť). Tento proces budeme reprezentovat pomocí Markovského rozhodovacího procesu.

Parametry mince, tedy pravděpodobnosti hodu panny, resp. orla, pro jednotlivé kombinace vstupu a paměti, nám přitom nebudou známy, popřípadě budou známy pouze částečně.

Naším cílem potom bude tento proces řídit, tedy najít optimální rozhodovací strategii (posloupnost funkcí přiřazujících v každém čase optimální vstup k posloupnosti předchozích vstupů a výstupů) tak, abychom v určitém časovém horizontu získali např. co největší počet hodů panny (úspěch), resp. minimalizovali počet hodů orla (ztráta).

Vzhledem k tomu, že neznáme přesně parametry naší mince, bude výběr optimálního vstupu o to složitější. V tomto případě hovoříme o tzv. rozhodování za nejistoty. V takovém případě je vhodné použít tzv. bayesovskou teorii. Výhodou použití této strategie je to, že samotný průběh řízení nejen minimalizuje očekávanou ztrátu, ale rovněž získává o procesu co nejvíce informací, na základě kterých minimalizuje budoucí ztráty. Takovýto způsob řízení se nazývá duální řízení [10], tj. řízení, které kombinuje samotné řízení s procesem učení.

Rozhodovací strategii za neurčitosti pak budeme hledat pomocí dynamického programování, které navrhl americký matematik Richard Bellman [9]. Jedná se o metodu, která s využitím zpětného chodu minimalizuje hodnotu očekávané ztrátové funkce. Přímá aplikace tohoto postupu je však bohužel i u poměrně jednoduchých systémů značně komplikovaná svou složitostí výpočtu. K řešení reálných úloh je proto vhodné použít aproximačních metod.

V průběhu práce si nejdříve v kapitole 2 nadefinujeme Markovský řetězec a Markovský rozhodovací proces. V kapitole 3 si přiblížíme některá rozdělení pravděpodobnosti, která budeme v tomto textu používat. V kapitole 4 se seznámíme s Bayesovskou teorií a dynamickým Bayesovským rozhodovacím procesem a nakonec v kapitole 5 aplikujeme naše znalosti na konkrétní příklad.

Poznámka 1 (poznámka ke značení) *Není-li dále v textu předem řečeno jinak, pak při značení:*

- *Nerozlišujeme hodnotu náhodné veličiny a samotnou náhodnou veličinu. Význam bude vždy zřejmý z kontextu.*
- *Nerozlišujeme náhodnou veličinu a náhodný vektor.*
- *Hustoty pravděpodobnosti jsou zapisovány bez explicitně uvedených náhodných veličin, ty jsou jednoznačně určeny argumenty dané hustoty.*
- *Zpravidla neuvádíme integrační obory. Integrály jsou vždy určité přes obor hodnot příslušné náhodné veličiny.*

2 Markovské řetězce

2.1 Markovský řetězec jako náhodný proces

Náhodný (stochastický) proces je zobrazení, které každé hodnotě $t \in T$, $T \in \mathbb{R}$, přiřadí náhodnou veličinu X_t , kde t má obvykle význam času. Čas t pak může být spojitý nebo diskrétní. V diskrétním případě je pak $T \in \mathbb{N}$.

Markovský řetězec je takový náhodný proces, který obvykle (a také v našem případě) pracuje s diskrétním časem - tzn., že je definován pouze v okamžicích tvořících (potenciálně nekonečnou) rostoucí posloupnost, mezi těmito okamžiky se pak nic neděje (popř. děje, ovšem v Markovském řetězci toto není zaznamenáno).

Definice 1 Mějme množinu stavů S , kde $S = \{1, 2, \dots, n\}$, $n \in \mathbb{N}$. Označme přechod z jednoho stavu do druhého jako krok. Markovský řetězec je náhodný proces, který splňuje tzv. Markovskou vlastnost, která říká, že pravděpodobnost přechodu ze stavu i , $i \in S$, do stavu j , $j \in S$, nezávisí na tom, ve kterých stavech se řetězec nacházel před výchozím stavem i , tedy

$$P(x_t = j | x_{t-1} = i, x_{t-2} = k_{t-2}, \dots, x_0 = k_0) = P(x_t = j | x_{t-1} = i),$$

pro libovolná $i, j, k_0, \dots, k_{t-2} \in S$.

Markovský řetězec nazveme homogenní, pokud pro něj platí, že pravděpodobnosti p_{ij} nezávisí na čase t .

Nehomogenními Markovskými řetězci, tedy takovými, jejichž pravděpodobnosti přechodu jsou závislé na čase, se v téhle práci nebudeme zabývat.

Poznámka 2 (Značení) Necht' se řetězec právě nachází ve stavu i ($i \in S$), pak pravděpodobnost toho, že řetězec v následujícím kroku přejde do stavu j ($j \in S$), značíme p_{ij} , tj.:

$$p_{ij} = P(x_t = j | x_{t-1} = i).$$

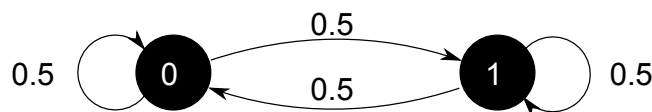
2.2 Přechodová matice

Pravděpodobnosti přechodu p_{ij} můžeme uspořádat do čtvercové stochastické matice. Takovou matici pak nazýváme přechodovou maticí příslušného Markovského řetězce.

Definice 2 Mějme matici $P = (p_{ij})$. Matice P se nazývá stochastická, platí-li:

$$(1.) \forall i, j : p_{ij} \in \langle 0, 1 \rangle$$

$$(2.) \forall i : \sum_j p_{ij} = 1$$



Obrázek 1: Markovský řetězec představující hod pravidelnou symetrickou mincí.

2.3 Absorpční řetězec

Absorpční řetězec je takový Markovský řetězec, který obsahuje alespoň jeden absorpční stav. Absorpční stav je pak takový stav i , pro který platí, že pravděpodobnost setrvání v tomto stavu je rovna 1, tedy

$$p_{ii} = P(x_t = i | x_{t-1} = i) = 1.$$

Protože platí

$$\sum_j p_{ij} = 1,$$

$$\sum_{j, j \neq i} p_{ij} = 1 - p_{ii},$$

$$\sum_{j, j \neq i} p_{ij} = 0,$$

$$\forall j, j \neq i : p_{ij} = 0,$$

tak se Markovský řetězec po dosažení absorpčního stavu i již z tohoto stavu do jiného nedostane.

2.4 Příklady Markovského řetězce

Příklad 1 Uvažujme hod pravidelnou symetrickou mincí, tzn. pravděpodobnost, že padne panna nebo orel, je stejná, tedy 50%. Pak náhodný proces představující hod touto mincí můžeme popsat následujícím Markovským řetězcem (viz také obrázek 1).

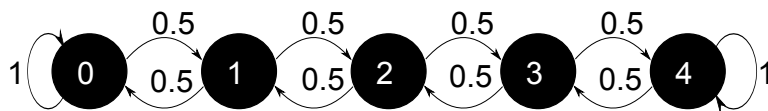
Nechť $t \in \mathbb{N}$ představuje pořadí pokusu hodu touto mincí a náhodná veličina X výsledek jejího hodu. Necht'

$$x_t = \begin{cases} 0, & \text{padne-li panna} \\ 1, & \text{padne-li orel} \end{cases}$$

A tedy

$$P(x_t = 1 | x_{t-1} = c) = P(x_t = 0 | x_{t-1} = c) = 0,5$$

pro libovolné $c \in \{0, 1\}$. Tedy pravděpodobnost $P(x_t | x_{t-1})$ nezávisí na x_{t-1} , to znamená, že x_t a x_{t-1} jsou nezávislé náhodné veličiny.



Obrázek 2: Markovský řetězec představující jednoduchou hru o body.

$$P(x_t = 1) = P(x_t = 0) = 0,5$$

I přesto však můžeme tento speciální (triviální) případ modelovat Markovským řetězcem.

Přechodová matice tohoto Markovského řetězce vypadá následovně (pro praktické účely indexujeme tuto matici od 0)

$$P = \begin{bmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{bmatrix}.$$

Příklad 2 Uvažujme jednoduchou hru mezi dvěma hráči. Na začátku této hry má každý hráč 2 body a hází si pravidelnou symetrickou mincí z předchozího příkladu. Pokud padne panna, pak vyhrává první hráč a získává jeden bod od svého soupeře, pokud padne orel, tak vyhrává hráč druhý jeden bod od prvního. Hra končí v případě, kdy jeden z hráčů nemá k dispozici žádný bod.

Pak můžeme tuto hru z pohledu jednoho z hráčů modelovat následujícím Markovským řetězcem (viz obrázek 2), kde množina stavů $S = \{0, 1, 2, 3, 4\}$, kde každý stav představuje počet bodů, které tento hráč vlastní.

Stav 0, resp. 4, je absorpční stav, po jeho dosažení hra končí prohrou, resp. vítězstvím, tohoto hráče.

Přechodová matice tohoto Markovského řetězce vypadá následovně (pro praktické účely indexujeme tuto matici od 0)

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

2.5 Markovský rozhodovací proces

Markovský rozhodovací proces je takový Markovský řetězec, který řídíme volbou akcí, jenž ovlivňují chování tohoto procesu. V každém časovém okamžiku zvolíme akci a na jejím základě proces reaguje.

Oproti Markovskému řetězci je v Markovském rozhodovacím procesu pravděpodobnost přechodu v čase t ze stavu i v předchozím čase $t - 1$ do stavu j tedy závislá nejen na stavu i (Markovská vlastnost), ale navíc také na zvolené akci a v čase t . Označme tuto pravděpodobnost $p_{a,i,j}$, pak tedy platí:

$$p_{a,i,j} = P(x_t = j | x_{t-1} = i, a_t = a),$$

kde $a \in \mathcal{A}_i$, kde \mathcal{A}_i je množina všech možných akcí daného Markovského rozhodovacího procesu nacházejícího se aktuálně ve stavu i . Zpravidla je však množina \mathcal{A}_i na stavu i nezávislá.

Příklad 3 (Jednoduchý Markovský rozhodovací proces) Uvažujme hod nesymetrickou mincí, u které pravděpodobnost hodu panny nebo orla je závislá na tom, v jaké poloze se mince nachází před provedením samotného hodu a na předchozím výsledku hodu touto mincí.

Nechť pro tuto minci platí:

$$x_t = \begin{cases} 0, & \text{padne-li panna} \\ 1, & \text{padne-li orel} \end{cases}$$

a necht'

$$a_t = \begin{cases} 0, & \text{je-li mince před hodem otočená pannou vzhůru,} \\ 1, & \text{je-li mince před hodem otočená orlem vzhůru.} \end{cases}$$

Nechť je dále pravděpodobnost hodu panny či orla závislá na tom, jak máme minci před hodem otočenou a jak dopadl předchozí hod touto mincí, např. takto:

$$\begin{aligned} P(x_t = 0 | x_{t-1} = 0, a_t = 0) &= 0,5, \\ P(x_t = 1 | x_{t-1} = 0, a_t = 0) &= 0,5, \\ P(x_t = 0 | x_{t-1} = 1, a_t = 0) &= 0,75, \\ P(x_t = 1 | x_{t-1} = 1, a_t = 0) &= 0,25, \\ P(x_t = 0 | x_{t-1} = 0, a_t = 1) &= 0,7, \\ P(x_t = 1 | x_{t-1} = 0, a_t = 1) &= 0,3, \\ P(x_t = 0 | x_{t-1} = 1, a_t = 1) &= 0,25, \\ P(x_t = 1 | x_{t-1} = 1, a_t = 1) &= 0,75. \end{aligned}$$

Tedy např. pravděpodobnost, že padne panna, je-li předchozí výsledek hodu orel a zvolíme-li akci otočení mince orlem vzhůru, je třikrát menší, než že padne orel, ale pokud bychom zvolili opačnou akci, tedy pokud bychom minci otočili před hodem pannou nahoru, byla by zase pravděpodobnost hodu panny třikrát vyšší, než pravděpodobnost hodu orla.

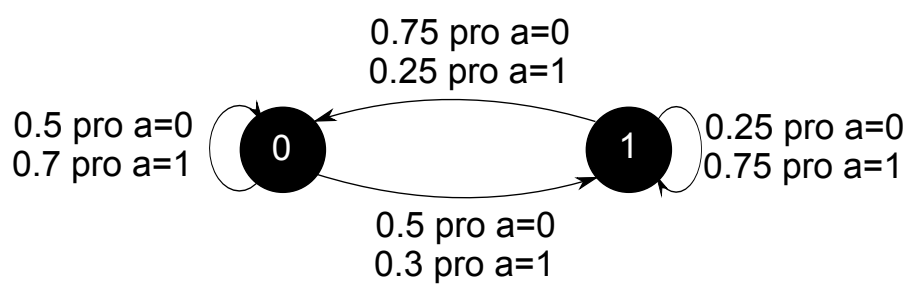
Pro tento příklad můžeme přechodové pravděpodobnosti zapsat formou dvojice stochastických matic, pro každou akci jednu. Matici P_0 pro pravděpodobnosti přechodů při zvolené akci 0

$$P_0 = \begin{bmatrix} 0,5 & 0,5 \\ 0,75 & 0,25 \end{bmatrix}$$

a matici P_1 pro akci 1

$$P_1 = \begin{bmatrix} 0,7 & 0,3 \\ 0,25 & 0,75 \end{bmatrix}.$$

Tento Markovský rozhodovací proces je znázorněn na obrázku 3.



Obrázek 3: Markovský rozhodovací proces znázorňující hod nesymetrickou mincí.

3 Vybraná rozdělení pravděpodobnosti

Rozdělení náhodné veličiny X je předpis charakterizující pravděpodobnost jevů, jenž lze touto náhodnou veličinou popsat. Pro diskrétní náhodnou veličinu je tímto předpisem obvykle pravděpodobnostní funkce, pro spojitou náhodnou veličinu je předpisem naopak distribuční funkce, popř. hustota pravděpodobnosti.

V této kapitole si představíme některá rozdělení pravděpodobnosti, kterými se budeme dále v této práci zabírat.

3.1 Alternativní rozdělení

Nechť je dána náhodná veličina X , která nabývá pouze dvou hodnot, 0 a 1. Obvykle tato náhodná veličina popisuje výsledek náhodného pokusu, při kterém sledujeme výskyt určitého náhodného jevu, přičemž náhodná veličina nabývá hodnoty 1, pokud daný náhodný jev nastane (většinou označováno jako úspěch), a naopak hodnoty 0, pokud nenastane.

Pravděpodobnostní funkce náhodné veličiny X s alternativním rozdělením s parametrem θ , kde θ je pravděpodobnost úspěchu, vypadá následovně

$$f(x) = \begin{cases} \theta, & \text{pro } x = 1, \\ 1 - \theta, & \text{pro } x = 0. \end{cases}$$

Vhodnější je však pracovat s pravděpodobnostní funkcí v následujícím tvaru

$$f(x) = \theta^x (1 - \theta)^{1-x}.$$

To, že má náhodná veličina X alternativní rozdělení s parametrem θ , značíme $X \sim \text{Alt}(\theta)$. Střední hodnota tohoto rozdělení je

$$\mu = \theta$$

a rozptyl

$$\sigma^2 = \theta(1 - \theta).$$

3.2 Binomické rozdělení

Náhodná veličina popisující četnost výskytu náhodného jevu v n nezávislých pokusech, v nichž má tento jev stále stejnou pravděpodobnost θ , má binomické rozdělení s parametry n, θ . To budeme značit $X \sim \text{Bi}(n, \theta)$.

Pravděpodobnostní funkce binomického rozdělení je

$$f(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$$

střední hodnota

$$\mu = n\theta,$$

a rozptyl

$$\sigma^2 = n\theta(1 - \theta).$$

Povšimněme si, že alternativní rozdělení je vlastně speciálním případem binomického rozdělení s parametrem $n = 1$.

3.3 Rovnoměrné rozdělení

Nechť je dán interval (a, b) , kde $-\infty < a < b < \infty$. Rovnoměrné rozdělení je rozdělení, jehož hustota pravděpodobnosti je konstantní na intervalu (a, b) a všude jinde je nulová. Tedy

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{pro } x \in (a, b) \\ 0, & \text{jinde.} \end{cases}$$

To, že má náhodná veličina X rovnoměrné rozdělení na intervalu (a, b) , značíme $X \sim R(a, b)$. Střední hodnota vypočteme jako

$$\mu = \frac{a+b}{2}$$

a rozptyl

$$\sigma^2 = \frac{(b-a)^2}{12}.$$

Diskrétní rovnoměrné rozdělení, které popisuje náhodnou veličinu nabývající n hodnot, pak přiřazuje všem těmto hodnotám stejnou pravděpodobnost $\frac{1}{n}$.

3.4 Beta rozdělení

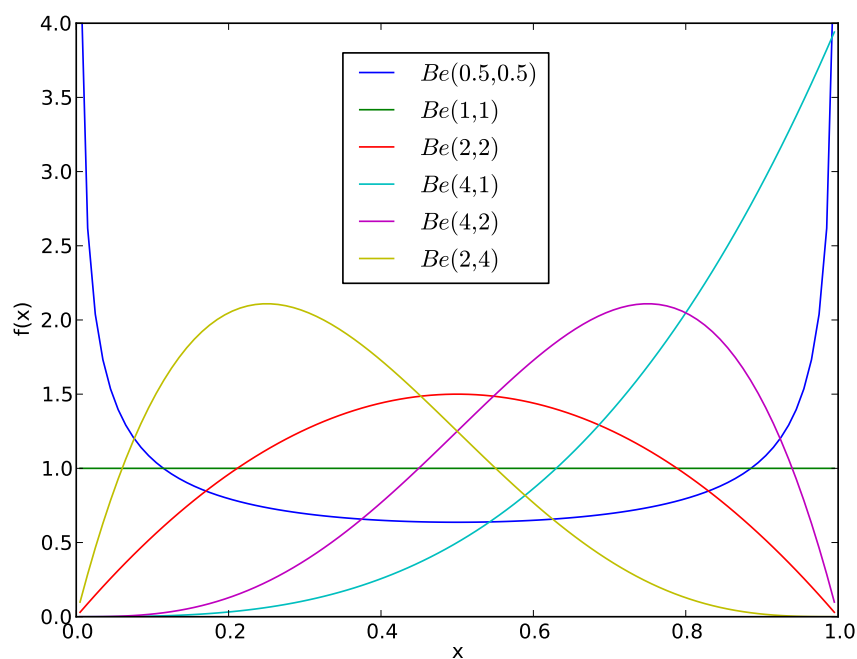
Náhodná veličina X (nabývající hodnot z intervalu $(0, 1)$) má beta rozdělení s parametry α, β , značíme $X \sim Be(\alpha, \beta)$, jestliže je její funkce hustoty pravděpodobnosti

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)},$$

kde

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt.$$

je tzv. beta funkce vystupující jako normalizační konstanta. Beta funkci $B(\alpha, \beta)$ můžeme zapsat pomocí podílu gama funkcí



Obrázek 4: Průběh funkce hustoty pravděpodobnosti beta rozdělení pro různé parametry.

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Pro gama funkci platí

$$\Gamma(t + 1) = t\Gamma(t).$$

Střední hodnota beta rozdělení s parametry α a β je

$$\mu = \frac{\alpha}{\alpha + \beta}$$

a rozptyl

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Na obrázku (4) je znázorněna hustota pravděpodobnosti pro beta rozdělení o různých kombinacích parametrů α a β .

Rozdělení pravděpodobnosti	Značení	$f(x)$	kde
Normální	$X \sim N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$
Poissonovo	$X \sim P(\theta)$	$\frac{\theta^x}{x!} e^{-\theta}$	$\theta \in \mathbb{R}^+$
Gama	$X \sim G(\alpha, \beta)$	$\frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$	$\alpha, \beta \in \mathbb{R}^+$

Tabulka 1: Další rozdělení pravděpodobnosti

3.5 Další rozdělení pravděpodobnosti

Další rozdělení pravděpodobnosti jsou uvedena v tabulce 1. Jsou v textu zmíněny jen okrajově, proto je zde uvedeno pouze jejich značení a hustota pravděpodobnosti, popř. pravděpodobnostní funkce. Více o těchto rozděleních pravděpodobnosti najdete např. v [8].

4 Bayesovské metody

Oproti klasickému statistickému přístupu je neznámý parametr θ rozdělení náhodné veličiny X při Bayesově přístupu chápán taktéž jako náhodná veličina. Základním nástrojem tohoto přístupu je Bayesova věta, jejíž autorem je Thomas Bayes (1701 - 1761).

Tato věta nám dává návod, jak ze znalosti pravděpodobnosti jevu A za podmínky B spočítat pravděpodobnost jevu B za podmínky A .

Věta 1 (Bayesova) *Necht' B_1, B_2, \dots, B_n tvoří úplný systém disjunktních jevů, tj.*

$$\bigcup_i B_i = \Omega,$$

kde Ω je jistý jev, a

$$\forall i, j, i \neq j : B_i \cap B_j = \emptyset,$$

a necht'

$$\forall i \in \{1, 2, \dots, n\} : P(B_i) > 0.$$

Necht' A je libovolný náhodný jev ze stejného pravděpodobnostního prostoru takový, že $P(A) > 0$. Pak pro všechna $i \in \{1, 2, \dots, n\}$ platí

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

4.1 Bayesův vzorec

Necht'

- θ je náhodná veličina s hustotou $f(\theta)$, $\theta \in \Theta$
- X je náhodná veličina s podmíněnou hustotou $f(x|\theta)$ při daném parametru θ ,

tedy

$$P(\theta \in B, x \in C) = \int_B \left(\int_C f(x|\theta) dx \right) f(\theta) d\theta,$$

kde B a C jsou libovolné měřitelné množiny.

Věta 2 (Bayesův vzorec) *Pro podmíněnou hustotu $f(\theta|x)$ náhodného vektoru θ při daném x platí*

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{\Theta} f(x|\theta)f(\theta)d\theta} \quad (1)$$

Hustotu $f(\theta)$ v (1) nazýváme *apriorní hustotou*, tato hustota vyjadřuje informaci o parametru θ ještě před samotnou realizací náhodné veličiny X . Naopak podmíněnou hustotu $f(\theta|x)$ nazýváme *hustotou aposteriorní*, která reprezentuje znalost o parametru θ .

Pak tedy k závěrům o parametru θ použijeme aposteriorní hustotu $f(\theta|x)$, ta v sobě zahrnuje jak apriorní informaci o parametru θ , tak i informaci plynoucí z realizace X .

Poznámka 3 Existují-li funkce $h_1(\theta, x)$ a $h_2(x)$ takové, že

$$f(\theta)f(x|\theta) = h_1(\theta, x)h_2(x),$$

pak aposteriorní hustotu $f(\theta|x)$ můžeme upravit do tvaru

$$f(\theta|x) = \frac{h_1(\theta, x)}{\int h_1(\theta, x)d\theta}$$

Důsledek 1 Z poznámky 3 plyne, že nahradíme-li v (1) věty 2 $f(\theta)$ funkcí $cf(\theta)$, kde $c \in \mathbb{R}$, $c > 0$, pak se aposteriorní hustota $f(\theta|x)$ nezmění.

Tohoto důsledku využijeme v kapitole 4.2.2.

Střední hodnotu náhodné veličiny X_i vypočteme jako

$$EX_i = E_\theta[E[X_i|\theta]] = \int \left(\int X_i f(x|\theta) dx \right) f(\theta) d\theta.$$

Na závěr této kapitoly ještě uvedme nepodmíněnou hustotu náhodné veličiny X ,

$$f(x) = \int f(x|\theta)f(\theta)d\theta.$$

4.2 Volba apriorního rozdělení

Jak jsme již zmínili v předchozí kapitole, k závěrům o parametru θ použijeme kromě náhodné veličiny X také informaci (byť neúplnou) o parametru θ nezávislou na realizaci X . Tuto informaci nazýváme *apriorní informace*.

Apriorní informace může mít buď objektivní (např. informace získaná z podobné úlohy), nebo subjektivní charakter (např. názor či zkušenost nějakého subjektu). Naše apriorní informace o parametru θ by pak měla odrážet apriorní hustota $f(\theta)$.

Chceme-li použít bayesovský přístup, řešíme problém volby apriorního rozdělení. Podobně jako při klasickém přístupu dáváme přednost práci s hustotami $f(\theta)$ určitého funkcionálního typu. Používáme např. tzv. systémy konjugovaných rozdělení (viz kapitola (4.2.1)) nebo typy rozdělení odpovídající tzv. principu neurčitosti (viz kapitola (4.2.2)), který používáme v případě, kdy nemáme o parametru θ informaci žádnou. V případě, kdy nemáme k dispozici žádnou apriorní informaci, můžeme použít také tzv. Jeffreysovu neinformativní hustotu (viz kapitola (4.2.3)).

4.2.1 Konjugované systémy hustot

Nechť $\mathbf{X} = (X_1, \dots, X_n)$ je náhodný výběr z rozdělení s hustotou $f(\mathbf{x}|\theta)$.

Definice 3 Nechť \mathcal{F} značí systém hustot $f(\mathbf{x}|\theta)$ pro různé parametry θ , tj.:

$$\mathcal{F} = \{f(\mathbf{x}|\theta), \theta \in \Theta\}.$$

Systém \mathcal{P} apriorních hustot $f(\theta)$ nazveme systémem konjugovaným se systémem hustot \mathcal{F} , pokud pro všechna $f \in \mathcal{F}$ a každé apriorní rozdělení z \mathcal{P} patří i aposteriorní hustota $\pi(\theta|\mathbf{x})$ jakožto hustota θ do systému \mathcal{P} .

Příkladem konjugovaného systému hustot může být např. množina všech možných rozdělení na Θ . To však není příliš praktické a vhodné je použít konjugovaný systém s konečným počtem rozdělení. Výhodou takového konjugovaného systému hustot pak je její jednoduchý přechod z apriorní na aposteriorní hustotu - a to pouhou změnou příslušných parametrů.

Konjugovaný systém hustot můžeme sestavit například pomocí metody využívající postačující statistiky.

Definice 4 Uvažujme statistický model $f(\mathbf{x}|\theta)$ náhodné veličiny X s hodnotami v \mathcal{X} . Nechť pro $T : \mathcal{X} \rightarrow \mathbb{R}^n, n \in \mathbb{N}$ platí

$$f(\mathbf{x}|T(\mathbf{X}) = t, \theta) = f(\mathbf{x}|T(\mathbf{X}) = t),$$

pak T nazýváme postačující statistikou modelu $f(\mathbf{x}|\theta)$.

Poznámka 4 Postačující statistika nese stejnou informaci o θ jako náhodná veličina X .

Důkaz: Nechť $t = T(\mathbf{x})$. Pak

$$\begin{aligned} f(\theta|\mathbf{x}) &= f(\theta|\mathbf{x}, t) = \frac{f(\mathbf{x}, t|\theta)f(\theta)}{\int f(\mathbf{x}, t|\theta)f(\theta)d\theta} = \frac{f(\mathbf{x}|t, \theta)f(t|\theta)f(\theta)}{\int f(\mathbf{x}|t, \theta)f(t|\theta)f(\theta)d\theta} = \\ &= \frac{f(\mathbf{x}|t)f(t|\theta)f(\theta)}{f(\mathbf{x}|t)\int f(t|\theta)f(\theta)d\theta} = \frac{f(t|\theta)f(\theta)}{\int f(t|\theta)f(\theta)d\theta} = f(\theta|t), \end{aligned}$$

kde první rovnost můžeme použít, protože postačující statistika nepřináší o θ žádnou informaci navíc oproti náhodné veličině X .

Tedy $f(\theta|\mathbf{x}) = f(\theta|t)$. \square

Příklad 4 (Postačující statistika alternativního rozdělení) Dokažme, že statistika

$$T_n(X_1, \dots, X_n) = \sum_{i=1}^n X_i$$

je postačující statistikou příslušící alternativnímu rozdělení.

Nechť náhodný výběr $\mathbf{X}_n = (X_1, \dots, X_n)'$ pochází z alternativního rozdělení s parametrem $\theta \in (0, 1)$, tj. $X_i \sim A(\theta)$. Nechť

$$S_n = \sum_{i=1}^n X_i.$$

Pak

$$S_n \sim Bi(n, \theta)$$

Nechť $\mathbf{x}_n = (x_1, \dots, x_n)'$ je realizace náhodného výběru. Uvažujme podmíněnou pravděpodobnost pro libovolně, ale pevně zvolené $t \in \mathbb{R}$:

$$P(X_1 = x_1, \dots, X_n = x_n | S_n = t).$$

Pro $\sum_{i=1}^n x_i \neq t$ je tato podmíněná pravděpodobnost rovna nule.

Nechť $\sum_{i=1}^n x_i = t$. Pak

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | S_n = t) &= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P_\theta(S_n = t)} = \\ &= \frac{\prod_{i=1}^n P(X_i = x_i)}{P(S = t)} = \\ &= \frac{\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \\ &= \frac{1}{\binom{n}{t}}. \end{aligned} \quad (2)$$

A tedy tato podmíněná pravděpodobnost nezávisí na θ , takže statistika $S_n = \sum_{i=1}^n X_i$ je postačující statistikou alternativního rozdělení. \square

Poznámka 5 Nechť $T_n(X_1, \dots, X_n)$ značí postačující statistiku příslušící systému hustot $\mathcal{F} = \{\prod_{i=1}^n f(x_i | \theta); \theta \in \Theta\}$. Předpokládejme, že existuje n_0 takové, že pro všechna $n \geq n_0$ existují nezáporné funkce g_n a h_n takové, že

$$\prod_{i=1}^n f(x_i | \theta) = g_n(T_n(x), \theta) h_n(x),$$

Označme S_n množinu všech bodů t , kterých může nabývat náhodný vektor $T_n(X)$.

$$S_n = \{t; t = T_n(X)\}.$$

Pak (dle [4]), za předpokladu

$$0 < \int g_n(t, \theta) d\theta < +\infty,$$

je systém hustot

$$\{f_{n,t}(\theta); t \in S_n, n \geq n_0\},$$

kde

$$f_{n,t}(\theta) = \frac{g_n(t, \theta)}{\int g_n(t, \theta) d\theta}, \quad (3)$$

systém konjugovaný s \mathcal{F} .

Příklad 5 (Konstrukce konjugovaného systému hustot) Mějme alternativní rozdělení s parametrem $\theta \in \langle 0, 1 \rangle$. Sestavme si pomocí metody pro konstrukci konjugovaného systému hustot využívající postačující statistiky konjugovaný systém hustot příslušící tomuto rozdělení.

Nechť pro postačující statistiku alternativního rozdělení

$$T_n(X_1, \dots, X_n) = \sum_{i=1}^n X_i$$

označíme S_n množinu všech možných hodnot, kterých může náhodná veličina T_n nabývat, tedy

$$S_n = \{1, 2, \dots, n\}.$$

Vzhledem k (1) je

$$f_n(y, \theta) = \theta^y (1 - \theta)^{n-y}, \quad y \in S_n,$$

$$h_n(x) = \prod_{i=1}^n \binom{1}{x_i}.$$

Odtud a z (3) plyne, že

$$f_{n,t}(\theta) = \frac{\theta^y (1 - \theta)^{n-y}}{B(y+1, n-y+1)}.$$

Pak konjugovaný systém hustot příslušící alternativnímu rozdělení je beta rozdělení s parametry $(y+1)$ a $(n-y+1)$, kde $y \in S$. Takovýto systém se nazývá přirozený konjugovaný systém. Často však pracujeme s bohatším systémem, v tomto případě by se jednalo o systém beta rozdělení s parametry α a β , kde $\alpha > 0$, $\beta > 0$, takový systém nazýváme obvyklý konjugovaný systém.

Je-li apriorní rozdělení beta rozdělení s parametry α, β , je aposteriorní rozdělení beta rozdělení s parametry $(\alpha + \sum_{i=1}^n X_i), (\beta + n - \sum_{i=1}^n X_i)$.

Příklady konjugovaných apriorních hustot pro některé další modely je zobrazeno v tabulce 2. V této tabulce je vždy parametr θ neznámý, všechny další parametry jsou pak známe.

Model	Rozdělení X	Konjugované rozdělení θ
Normální	$N(\theta, \sigma^2)$	$N(\mu, \epsilon^2)$
Normální	$N(\mu, \frac{1}{\theta})$	$G(\alpha, \beta)$
Poissonovo	$P(\theta)$	$G(\alpha, \beta)$
Gama	$G(v, \theta)$	$G(\alpha, \beta)$
Binomické	$Bi(n, \theta)$	$B(\alpha, \beta)$

Tabulka 2: Konjugované apriorní hustoty některých modelů

4.2.2 Princip neurčitosti

Necht' $f(x|\theta)$ je podmíněná hustota náhodné veličiny x při dané hodnotě parametru $\theta \in \Theta$. Pokládáme-li θ za náhodnou veličinu, o níž víme jen to, že $\theta \in \Theta$, vzniká problém jak volit apriorní rozdělení.

Nastává možnost volit za apriorní rozdělení θ rovnoměrné rozdělení na Θ . Příslušnou hustotu budeme značit $f_0(\theta)$. Tento přístup je (např. v [4]) označován jako princip neurčitosti.

Pokud je Θ nanejvýš spočetná množina, pak jde o hustotu vzhledem k číselní míře. Pokud má Θ kladnou Lebesgueovu míru, je f_0 hustotou vzhledem k Lebesgueově míře. V obou těchto případech je f_0 identicky rovna kladné konstantě. Pro účely výpočtu aposteriorního rozdělení ji můžeme položit rovno 1, viz důsledek poznámky 3 (kdy c položíme rovno $\frac{1}{f_0}$).

Nakonec, je-li Θ nekonečná spočetná, popř. je-li její Lebesgueova míra nekonečná, volíme za hustotu $f_0(\theta)$ tzv. nevlastní hustotu. Nevlastní hustota je definovaná jako nezáporná měřitelná funkce (nemusí být integrovatelná).

4.2.3 Jeffreysova hustota

V případě, kdy nemáme k dispozici žádnou vhodnou apriorní informaci, můžeme využít tzv. neinformativních apriorních rozdělení. Jednou z nejčastěji používanou je Jeffreysova hustota.

Uvažujme statistický model $f(x|\theta)$, $\theta \in \Theta$. Fisherova informace $I(\theta)$ pro $\Theta \in \mathbb{R}$ je definována vztahem

$$I(\theta) = E \left[\left(\frac{\partial \ln f(x|\theta)}{\partial \theta} \right)^2 \middle| \theta \right]$$

$I(\theta)$ vyjadřuje množství informace, které o parametru θ přinese pozorování x . Tato informace je závislá na θ .

Jeffreysova apriorní hustota pro model $f(x|\theta)$ je rovna funkci $(I(\theta))^{-\frac{1}{2}}$. Tato hustota bývá často nevlastní (tedy nemá konečný integrál).

Pro alternativní rozdělení je Jeffreysova hustota beta rozdělení s parametry $1/2$ a $1/2$.

4.3 Statistické rozhodovací funkce

Úloha rozhodování nastává v případě, kdy se musíme rozhodnout mezi alespoň dvěma různými možnostmi, které označujeme jako akce. Kritériem výběru daného rozhodnutí, které přijmeme, je většinou založeno na posouzení výsledků všech uvažovaných akcí. Rozhodování pak chápeme jako nějaký druh optimalizace.

Pokud nejsme schopni určit výsledky všech uvažovaných akcí přesně, je výběr nejlepší akce o to komplikovanější. Hovoříme pak o tzv. rozhodování za nejistoty. Nejistotou pak může být již zmiňovaná neúplná znalost o parametru θ . Pro tento účel je vhodné použít právě Bayesovský přístup.

4.3.1 Ztrátová, rozhodovací a riziková funkce

Nechť

- X je náhodná veličina s hustotou $f(x|\theta)$ nabývající hodnot v \mathcal{X} ,
- $\theta \in \Theta$ je parametr, Θ je neprázdná borelovská podmnožina \mathbb{R}^k .
- \mathcal{A} značí množinu možných rozhodnutí,
- $a \in \mathcal{A}$ je akce (rozhodnutí).

Nyní označme $L(\theta, a)$ ztrátovou funkci,

$$L : \Theta \times \mathcal{A} \rightarrow [C, +\infty], \quad (4)$$

kde $C \in \mathbb{R}$. Ztrátová funkce číselně vyjadřuje ztrátu, kterou utrpíme, jestliže skutečná hodnota parametru je θ a jestliže přijmeme rozhodnutí a .

Zobrazení $\delta, \delta : \mathcal{X} \rightarrow \mathcal{A}$, nazveme rozhodovací funkcí. Hodnota $\delta(x), x \in \mathcal{X}$, představuje rozhodnutí (závěr) pro $X = x$.

Střední ztrátu způsobenou volbou rozhodovací funkce δ , je-li skutečná hodnota parametru rovna θ , nazveme rizikem, označíme $R(\theta, \delta)$ a definujeme jako

$$R(\theta, \delta) = E [L(\theta, \delta(x)) | \theta] = \int L(\theta, \delta(x)) f(x|\theta) dx. \quad (5)$$

Nechť Δ značí množinu takových rozhodovacích funkcí δ , pro něž platí

$$\forall \theta \in \Theta : R(\theta, \delta) < +\infty.$$

Pak funkci $R : \Theta \times \Delta \rightarrow \mathbb{R}$, definovanou vztahem (5), nazýváme rizikovou funkcí.

Statistickým rozhodovacím problémem budeme rozumět trojici (Θ, Δ, R) .

4.3.2 Uspořádání a ekvivalence na množině Δ

Nechť $\delta_1, \delta_2 \in \Delta$. Pak platí:

Rozhodovací funkce δ_1 je R -lepší (tj. lepší vzhledem k rizikové funkci R) než rozhodovací funkce δ_2 , jestliže

$$R(\theta, \delta_1) \leq R(\theta, \delta_2), \quad \forall \theta \in \Theta$$

a jestliže alespoň pro jedno θ platí tato nerovnost ostrá.

Rozhodovací funkce δ_1 je R -ekvivalentní rozhodovací funkci δ_2 , jestliže

$$R(\theta, \delta_1) = R(\theta, \delta_2), \quad \forall \theta \in \Theta.$$

Optimální rozhodovací funkcí by zřejmě byla funkce $\delta^* \in \Delta$ taková, že

$$\min_{\delta \in \Delta} R(\theta, \delta) = R(\theta, \delta^*), \quad \forall \theta \in \Theta.$$

Taková rozhodovací funkce však existuje jen výjimečně.

4.3.3 Bayesovské riziko

Při bayesovském přístupu předpokládáme, že je parametr θ náhodná veličina s hustotou $f(\theta)$.

Definice 5 Pro $f(\theta)$ a $\delta \in \Delta$ definujeme Bayesovskou rizikovou funkci jako

$$\rho(f, \delta) = E[R(\theta, \delta)] = \int \left(\int L(\theta, \delta(x)) f(x|\theta) dx \right) f(\theta) d\theta.$$

Optimální bayesovskou rozhodovací funkci $\delta^*(f)$ definujeme jako

$$\min_{\delta \in \Delta} \rho(f, \delta) = \rho(f, \delta^*) \tag{6}$$

Označme $\rho^*(f) = \rho(f, \delta^*)$. Číslo $\rho^*(f)$ nazýváme bayesovské riziko.

Bayesovská rozhodovací funkce δ^* (6) závisí na volbě ztrátové funkce a apriorním rozdělení parametru θ .

Mějme dvě ztrátové funkce L_1 a L_2 , mezi nimiž platí vztah

$$L_1(\theta, d) = aL_2(\theta, d) + b$$

pro všechna $(\theta, d) \in \Theta \times \mathcal{D}$ a nějaké $a > 0, b \in \mathbb{R}$. Pak bayesovské rozhodovací funkce odpovídající L_1 a L_2 jsou shodné, což vzhledem k předpokladu omezenosti ztrátové funkce zdola (viz (4)) vede k tomu, že bez újmy na obecnosti můžeme předpokládat nezápornost ztrátové funkce.

4.3.4 Princip minimaxu

Jiné řešení rozhodovacího problému poskytuje např. princip minimaxu, při kterém za optimální rozhodovací funkci budeme považovat rozhodovací funkci δ^0 s vlastností

$$\min_{\delta \in \Delta} \max_{\theta \in \Theta} R(\theta, \delta) = \max_{\theta \in \Theta} R(\theta, \delta^0).$$

Rozhodujeme se tedy pro alternativu, která minimalizuje naše nejvyšší možné náklady. Princip minimaxu používáme např. v případě, kdy o parametru θ víme jen to, že $\theta \in \Theta$.

4.4 Dynamický Bayesovský rozhodovací proces

Obdobný přístup volby optimálního rozhodnutí jako v předchozí kapitole, tedy minimalizace očekávané ztráty, lze aplikovat i na dynamické rozhodovací úlohy, tedy úlohy, kdy se rozhodnutí generuje sekvenčně v čase.

Mějme systém, který je popsán posloupností náhodných veličin $(a_t, \theta_t, x_t)_{t \in t^*}$, kde

- t^* je rozhodovací horizont, $t^* = \{1, 2, \dots, T\}$, kde $T \in \mathbb{N}$, $T > 1$, představuje čas, do kterého je rozhodování prováděno,
- θ_t je parametr ovlivňující chování systému (pro daný čas t),
- akce a_t rozhodnutí učiněné námi v rámci rozhodovacího problému (v daném čase t)
- a x_t značí výstup, tedy samotnou realizaci (v čase t).

Platí, že pro daný časový okamžik t nejprve zvolíme akci a_t , následně se generuje nový parametr θ_t a až poté je v závislosti na nich realizován výstup x_t .

Poznámka 6 (Značení) *Nechť*

- $x^{t_1:t_2}$ značí posloupnost náhodných veličin $(x_{t_1}, \dots, x_{t_2})$ pro dané $t_1, t_2 \in t^*$, obdobně pro akce a_t , $a^{t_1:t_2} := (a_{t_1}, \dots, a_{t_2})$,
- data $d_t := (a_t, x_t)$ značí vektor akcí a_t a výstupů x_t pro daný časový okamžik t ,
- *trajektorie* systému je posloupnost realizací všech uvažovaných veličin od začátku do konce daného rozhodovacího horizontu, tj. $(a^{1:T}, \theta^{1:T}, x^{1:T})$, dle našeho značení tedy $(a^{1:T}, \theta^{1:T}, x^{1:T}) := (a_1, \theta_1, x_1, \dots, a_T, \theta_T, x_T)$,
- a_t nabývá hodnot v \mathcal{A}_t , tedy $a_t \in \mathcal{A}_t$, pak $a^{1:T} \in \mathcal{A}^{1:T} := \mathcal{A}_1 \times \dots \times \mathcal{A}_T$, obdobně $x^{1:T} \in \mathcal{X}^{1:T}$ a $\theta^{1:T} \in \Theta^{1:T}$

Abychom byli schopni popsat daný systém, potřebujeme jakožto jeho uživatel zvolit následující distribuční funkce, jež představují naši vstupní informaci reprezentující znalost o systému:

1. Model pozorování, což je posloupnost podmíněných distribučních funkcí

$$(f(x_t|a_t, d^{1:t-1}, \theta^{1:t-1}))_{t \in t^*},$$

které modelují závislost výstupní proměnné na všech předešlých veličinách. Předpokládáme, že je tento model zvolen tak, že pro aktuální akci a_t a předcházející data $d^{1:t-1}$ je výstupní proměnná x_t závislá pouze na aktuálním parametru θ_t , tedy dostáváme

$$(f(x_t|a_t, d^{1:t-1}, \theta_t))_{t \in t^*}. \quad (7)$$

2. Model vývoje v čase

$$(f(\theta_t|a_t, d^{1:t-1}, \theta^{1:t-1}))_{t \in t^*}$$

modeluje vývoj neznámých parametrů. Obdobně jako u modelu pozorování předpokládáme, že pro danou akci a_t a předcházející data $d^{1:t-1}$ je parametr θ_t závislý již pouze na předešlém parametru θ_{t-1} . Model tedy můžeme zapsat jako

$$(f(\theta_t|a_t, d^{1:t-1}, \theta_t))_{t \in t^*}.$$

3. Apriorní rozdělení

$$f(\theta_0),$$

jenž představuje naši počáteční znalost o neznámém parametru θ_0 .

Cílem rozhodování je pak uspořádat všechny možné trajektorie tak, abychom byli schopni některou trajektorii upřednostnit před jinou. K tomu nám poslouží ztrátová funkce

$$L : \mathcal{A}^{1:T} \times \mathcal{X}^{1:T} \times \Theta^{1:T} \rightarrow [C, +\infty], \quad (8)$$

pro nějaké $C \in \mathbb{R}$. Pomocí ztrátové funkce přiřadíme každé trajektorii hodnotu, čím menší hodnotu ztrátové funkce daná trajektorie má, tím spíše ji upřednostníme.

Na základě vstupních informací (modelu a apriorního rozdělení) a ztrátové funkce, která reprezentuje naše cíle, pak hledáme rozhodovací strategii.

Rozhodovací strategií je pak posloupnost zobrazení (indexovaných časem t) přiřazující předcházejícím datům $d^{1:t-1}$ konkrétní hodnotu akce a_t (nebo distribuční funkci akce a_t). Rozhodovací strategii představuje posloupnost podmíněných distribučních funkcí

$$(f(a_t|d^{1:t-1}))_{t \in t^*}.$$

Jednotlivé podmíněné distribuční funkce $f(a_t|d^{1:t-1})$ nazýváme rozhodovacím pravidlem.

Rozhodovací strategie tedy generuje akce a_t v závislosti na předchozích datech $d^{1:t-1}$. Z toho plyne, že akce a_t a parametry $\theta^{1:t-1}$ jsou podmíněně nezávislé při daných datech $d^{1:t-1}$ (tzv. přirozené podmínky řízení [6]), tj.

$$f(a_t|d^{1:t-1}, \theta^{1:t-1}) = f(a_t|d^{1:t-1}). \quad (9)$$

Přímým důsledkem (9) je i podmíněná nezávislost

$$f(\theta_{t-1}|a_t, d^{1:t-1}) = f(\theta_{t-1}|d^{1:t-1}).$$

Dle kapitoly 4.3.2 pak hledáme optimální rozhodovací strategii

$$^OR := (^Of(a_t|d^{1:t-1}))_{t \in t^*},$$

jako strategii minimalizující očekávanou ztrátu

$$E [L(a^{1:T}, x^{1:T}, \theta^{1:T})]$$

vzhledem ke sdružené hustotě pravděpodobnosti na množině všech možných trajektorií

$$\begin{aligned} f(a^{1:t}, x^{1:t}, \theta^{1:t}) &= \\ &= \int \prod_{t=1}^{t^*} f(x_t|a_t, d^{1:t-1}, \theta_t) f(\theta_t|a_t, d^{1:t-1}, \theta_{t-1}) f(a_t|d^{1:t-1}) f(\theta_0) d\theta_0. \end{aligned} \quad (10)$$

Postup optimalizace může být chápán tak, že se skládá ze dvou vzájemně propojených částí - fáze učení a návrhu samotné optimální strategie. Cílem učení je získat informace o neznámých parametrech θ_t z již dostupných dat. Při návrhu rozhodovací strategie pak konstruuje rozhodovací strategie s ohledem na výsledky získané právě z fáze učení.

Učení

V libovolném čase t je veškerá naše dostupná znalost o neznámých parametrech (počáteční i získaná z dat $d^{1:t}$) zastoupena podmíněnou distribuční funkcí

$$f(\theta_t|d^{1:t}).$$

Její vyčíslení je cílem fáze učení.

Výpočet provedeme rekurzivně, přičemž rekurzi začneme v čase $t = 0$ za použití apriorní hustoty $f(\theta_0)$.

1. Časová aktualizace

$$\begin{aligned} f(\theta_t|a_t, d^{1:t-1}) &= \int f(\theta_t|a_t, d^{1:t-1}, \theta_{t-1}) f(\theta_{t-1}|a_t, d^{1:t-1}) d\theta_{t-1} = \\ &= \int f(\theta_t|a_t, d^{1:t-1}, \theta_{t-1}) f(\theta_{t-1}|d^{1:t-1}) d\theta_{t-1} \end{aligned} \quad (11)$$

2. Datová aktualizace

$$f(\theta_t|d^{1:t}) = \frac{f(x_t|a_t, d^{1:t-1}, \theta_t)f(\theta_t|a_t, d^{1:t-1})}{f(x_t|a_t, d^{1:t-1})}, \quad (12)$$

kde

$$f(x_t|a_t, d^{1:t-1}) = \int f(x_t|a_t, d^{1:t-1}, \theta_t)f(\theta_t|a_t, d^{1:t-1})d\theta_t. \quad (13)$$

Podmíněnou distribuční funkci $f(x_t|a_t, d^{1:t-1})$ z poslední rovnice obecně nazýváme výstupním modelem systému. Tato distribuční funkce predikuje chování systému v daném čase pro konkrétní historii dat a akci v tomto čase zvolenou.

Návrh optimální strategie

Optimální strategie, tj. minimalizace (8), je konstruována rekurzivně pomocí tzv. *Bellmanových funkcí* $\mathcal{V}_t : (d^{1:t})^* \rightarrow \mathbb{R}$,

$$\mathcal{V}_t(d^{1:t}) = \min_{f(a_t|d^{1:t})} E [\mathcal{V}_{t+1}(d^{1:t})|d^{1:t-1}]. \quad (14)$$

Rekurze je prováděna zpětným způsobem, začíná v čase $t = T$ a počáteční funkce \mathcal{V}_{T+1} je definována jako

$$\mathcal{V}_{T+1}(d^{1:T}) = E [L(a^{1:T}, x^{1:T}, \theta^{1:T})|d^{1:T}]. \quad (15)$$

K jejímu vyčíslení potřebujeme podmíněnou hustotu $f(\theta^{1:T}|d^{1:T})$ danou následujícím rekurzivním vztahem

$$f(\theta^{1:t}|d^{1:t}) = \frac{f(x_t|a_t, d^{1:t-1}, \theta_t)f(\theta_t|a_t, d^{1:t-1}, \theta_{t-1})}{f(x_t|a_t, d^{1:t-1})}f(\theta^{1:t-1}|d^{1:t-1}),$$

kde $f(x_t|a_t, d^{1:t-1})$ je dáno vztahem (12). Rekurse začíná s apriorním rozdělením $f(\theta_0)$.

Očekávanou hodnotu ve vztahu (14) počítáme vzhledem k

$$f(d_t|d^{1:t-1}) = f(x_t|a_t, d^{1:t-1})f(a_t|d^{1:t-1}),$$

určenou výstupním modelem systému z fáze učení (13) a příslušným rozhodovacím pravidlem $f(a_t|d^{1:t-1})$. Optimální strategie OR pak spočívá v minimalizaci argumentů (14), tj.

$$^Of(a_t|d^{1:t-1}) \in \operatorname{argmin}_{f(a_t|d^{1:t-1})} E [\mathcal{V}_{t+1}(d^{1:t})|d^{1:t-1}]. \quad (16)$$

$^Of(a_t|d^{1:t-1})$ z této rovnice nejsou určeny jednoznačně. Nicméně, vzhledem k tomu, že všechny minimalizace v (16) vedou ke stejné očekávané ztrátě, můžeme rozhodovací pravidla $^Of(a_t|d^{1:t-1})$ volit libovolně. Protože $E [\mathcal{V}_{t+1}(d^{1:t})|d^{1:t-1}]$ v (14) závisí lineárně na $f(a_t|d^{1:t-1})$, je zřejmé, že $^Of(a_t|d^{1:t-1})$ je jakákoliv distribuční funkce, která přiřazuje množině

$$\operatorname{argmin}_{a_t \in a^*} E [\mathcal{V}_{t+1}(d^{1:t})|a_t, d^{1:t-1}]$$

pravděpodobnost 1 a optimální rozhodovací strategie může být vždy vybrána jako nenáhodná.

5 Formulace problému

V této kapitole aplikujeme dynamické rozhodování za neurčitosti popsané v kapitole 4.4 na konkrétní proces. Tímto procesem bude hod mincí reprezentovaný pomocí Markovského rozhodovacího procesu (viz kapitola 2.5).

5.1 Model

Mějme dán následující diskrétní náhodný proces s pamětí a řídicími vstupy zaznamenávající opakované hody nesymetrickou mincí, pro který platí, že výsledek hodu (**výstup**) závisí na

- poloze mince před hodem (**akce, vstup**)
- a na výsledku předchozího hodu (**paměť**).

Skutečné parametry mince, tj. pravděpodobnosti hodu panny (resp. orla) pro jednotlivé kombinace vstupu a paměti, nám přitom nejsou (úplně) známy. Poznamenejme navíc, že jednotlivé pravděpodobnosti výsledku hodu mince jsou závislé pouze na vstupu a paměti, tedy jsou např. nezávislé na čase.

Označme strany mince následovně:

- panna =: 0,
- orel =: 1.

Nechť dále $t, t \in \{1, 2, \dots, T\}$, kde $T \in \mathbb{N}, T > 1$, je diskrétní čas našeho procesu. Pak $x_t, x_t \in \{0, 1\}$, značí výstup v čase t a $a_t, a_t \in \{0, 1\}$, zase vstup v čase t . Označme $d_t := (a_t, x_t)$ vektor vstupu a výstupu pro daný čas t .

Označme pravděpodobnost

$$P(x_t = i | x_{t-1} = j, a_t = k) =: \theta_{i|jk}.$$

Tj. $\theta_{i|jk}$ značí pravděpodobnost, že v čase t padne i za podmínky, že v předchozím čase $t-1$ padlo j , přičemž jsme zvolili akci k .

Pak $\forall j, k$ platí

$$\theta_{0|jk} + \theta_{1|jk} = 1.$$

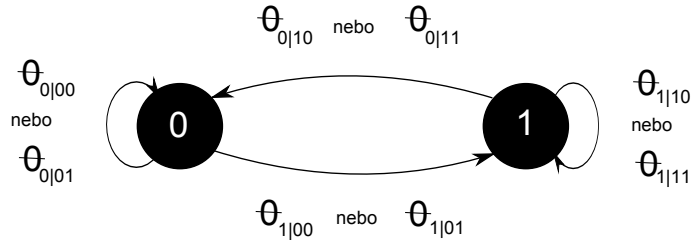
Hod touto mincí představuje Markovský rozhodovací proces, viz obrázek (5).

Označme si nově a dále pro jednoduchost použijeme pravděpodobnost hodu 1 jako

$$\theta_{jk} := \theta_{1|jk}$$

a vektor těchto pravděpodobností

$$\theta := (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}).$$



Obrázek 5: Náš uvažovaný model jako Markovský rozhodovací proces, kde pravděpodobnosti přechodu závisí na zvolené akci.

5.2 Rozhodovací úloha

Necht' v našem případě hod panny znamená úspěch a naopak. Naší cílem je navrhnout takovou strategii volby vstupů (tedy polohy mince před hodem), která povede právě k největšímu možnému počtu úspěchů, tedy hodů panny, přičemž přesně neznáme parametry této mince. Rozhodovat se budeme nejlépe až do konce časového horizontu T , tedy na T kroků (hodů) dopředu.

Řešíme tedy úlohu, ve které hledáme rozhodovací strategii

$$(f(a_t | d^{1:t-1}))_{t=1}^T$$

takovou, aby očekávaná ztráta

$$E(L(d^{1:T}))$$

byla co nejmenší, přičemž

$$L(d^{1:T}) = \sum_{t=1}^T l(x_t) \quad (17)$$

je tzv. aditivní ztrátová funkce, kde

$$l(x_t) = x_t.$$

Tzn., že ztráta v daném hodu je závislá jen na jeho výsledku a celková ztráta je rovna součtu ztrát v každém hodu, tedy hodíme-li pannu, ztráta je nulová a celková ztráta se nezvětší, hodíme-li orla, je rovna 1 a celková ztráta se zvětší o 1.

Poznamenejme také, že naše ztrátová funkce (17) je jednodušší než ztrátová funkce (8), protože je závislá pouze na datech $d^{1:T}$, nikoliv na neznámých parametrech θ .

Tuto úlohu budeme řešit jako dynamický Bayesovský rozhodovací proces (viz kapitola (4.4)).

5.3 Aplikace vlastností modelu na rozhodovací úlohu

Parametry mince jsou *nezávislé na čase*, platí tedy, že

$$f(\theta_t | a_t, d^{1:t-1}, \theta_{t-1}) = \delta(\theta_t - \theta_{t-1}),$$

kde δ je tzv. Diracova δ -funkce a tedy

$$\forall t \in \{1, \dots, T\} : P(\theta_t = \theta_0) = 1.$$

Protože parametry mince nezávisí na čase, můžeme použít následující značení

$$\theta := \theta_t.$$

Tímto nám vypadává krok časové aktualizace (11) z fáze učení a datová aktualizace (12) se zjednoduší (při použití 9) na

$$\begin{aligned} f(\theta | d^{1:t}) &= \frac{f(d_t | d^{1:t-1}, \theta) f(\theta | d^{1:t-1})}{f(d_t | d^{1:t-1})} = \\ &= \frac{f(x_t | a_t, d^{1:t-1}, \theta) f(a_t | d^{1:t-1}, \theta) f(\theta | d^{1:t-1})}{\int f(x_t | a_t, d^{1:t-1}, \theta) f(a_t | d^{1:t-1}, \theta) f(\theta | d^{1:t-1}) d\theta} = \\ &= \frac{f(x_t | a_t, d^{1:t-1}, \theta) f(a_t | d^{1:t-1}) f(\theta | d^{1:t-1})}{f(a_t | d^{1:t-1}) \int f(x_t | a_t, d^{1:t-1}, \theta) f(\theta | d^{1:t-1}) d\theta} = \\ &= \frac{f(x_t | a_t, d^{1:t-1}, \theta) f(\theta | d^{1:t-1})}{\int f(x_t | a_t, d^{1:t-1}, \theta) f(\theta | d^{1:t-1}) d\theta}, \end{aligned} \tag{18}$$

kde $f(\theta | d^{1:t-1})$ je aposteriorní rozdělení v čase $t - 1$.

Podmíněné hustoty *modelu pozorování* (7) naší mince splňují *Markovskou vlastnost*, tedy

$$f(x_t | a_t, d^{1:t-1}, \theta) = f(x_t | a_t, x_{t-1}, \theta).$$

Pro zjednodušení výpočtu volíme *konjugovanou apriorní hustotu*. Protože můžeme předpokládat nezávislost neznámých parametrů $\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}$, budeme v našem případě za apriorní rozdělení brát součin beta rozdělení (více viz příklad 5 z kapitoly 4.2.1).

Pro parametr $\theta = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})'$, kde

$$\theta_{jk} = P(x_t = 1 | x_{t-1} = j, a_t = k) \sim Be(\alpha_{jk}, \beta_{jk})$$

je apriorní hustota

$$f(\theta) = \prod_{jk} \frac{\theta_{jk}^{\alpha_{jk}-1} (1 - \theta_{jk})^{\beta_{jk}-1}}{B(\alpha_{jk}, \beta_{jk})}.$$

Pak pro konkrétní trajektorii $d^{1:t}$ je aposteriorní hustota

$$\begin{aligned} f(\theta | d^{1:t}) &\propto \left(\prod_{\tau=1}^t f(x_\tau | a_\tau, x_{\tau-1}, \theta) \right) f(\theta) = \\ &= \left[\prod_{\tau=1}^t \left(\prod_{j,k} \theta_{jk}^{\delta(x_\tau, 1) \delta(a_\tau, k) \delta(x_{\tau-1}, j)} (1 - \theta_{jk})^{\delta(x_\tau, 0) \delta(a_\tau, k) \delta(x_{\tau-1}, j)} \right) \right] f(\theta) = \\ &= \left(\prod_{j,k} \theta_{jk}^{\sum_{\tau=1}^t \delta(x_\tau, 1) \delta(a_\tau, k) \delta(x_{\tau-1}, j)} (1 - \theta_{jk})^{\sum_{\tau=1}^t \delta(x_\tau, 0) \delta(a_\tau, k) \delta(x_{\tau-1}, j)} \right) f(\theta). \end{aligned}$$

Označíme-li

$$A_{jk}(t) = \sum_{\tau=1}^t \delta(x_\tau, 1) \delta(a_\tau, k) \delta(x_{\tau-1}, j)$$

a

$$B_{jk}(t) = \sum_{\tau=1}^t \delta(x_\tau, 0) \delta(a_\tau, k) \delta(x_{\tau-1}, j),$$

pak

$$f(\theta | d^{1:t}) = \prod_{jk} \frac{\theta_{jk}^{\alpha_{jk} + A_{jk}(t) - 1} (1 - \theta_{jk})^{\beta_{jk} + B_{jk}(t) - 1}}{B(\alpha_{jk} + A_{jk}(t), \beta_{jk} + B_{jk}(t))}, \quad (19)$$

tedy

$$\theta_{jk} | d^{1:t} \sim Be(\alpha_{jk} + A_{jk}(t), \beta_{jk} + B_{jk}(t)),$$

kde $A_{jk}(t)$ je počet přeskoků v dané trajektorii $d^{1:t}$ z $x_{\tau-1} = j$ při $a_\tau = k$ do $x_\tau = 1$ pro všechna $\tau \in \{1, \dots, t\}$ a $B_{jk}(t)$ počet těchto přeskoků do $x_\tau = 0$.

Obecně může ztrátová funkce záviset na akcích a_t , pozorováních x_t a neznámých parametrech θ_t . V našem případě však není ztrátová funkce (17) na neznámých parametrech závislá. To znamená, že se jinak složité vyčíslení *Bellmanovy funkce* (15) v čase $T + 1$ zjednodušuje na

$$\mathcal{V}_{T+1}(d^{1:T}) = L(d^{1:T}).$$

Nakonec využijeme ještě toho, že používáme *aditivní ztrátovou funkci* (17). Pak pro Bellmanovy funkce platí

$$\begin{aligned} \mathcal{V}_{t+1}(d^{1:t}) &= \min_{(f(a_{t'}|d^{1:t'-1}))_{t'=t}^T} E[L(d^{1:T})|d^{1:t}] = \\ &= \min_{(f(a_{t'}|d^{1:t'-1}))_{t'=t}^T} E\left[\sum_{\tau=1}^T x_{\tau} \middle| d^{1:t}\right] = \\ &= \min_{(f(a_{t'}|d^{1:t'-1}))_{t'=t}^T} E\left[\sum_{\tau=1}^t x_{\tau} + \sum_{\tau=t+1}^T x_{\tau} \middle| d^{1:t}\right] = \\ &= \sum_{\tau=1}^t x_{\tau} + \min_{(f(a_{t'}|d^{1:t'-1}))_{t'=t}^T} E\left[\sum_{\tau=t+1}^T x_{\tau} \middle| d^{1:t}\right], \end{aligned}$$

protože suma

$$\sum_{\tau=1}^t x_{\tau}$$

závisí pouze na datech $d^{1:t}$ a nezávisí na znáhodněných rozhodnutích

$$\left(f(a_{t'}|d^{1:t'-1})\right)_{t'=t}^T.$$

Stačí nám proto definovat Bellmanovy funkce jako minimum očekávaného přírůstku aditivní ztrátové funkce, tedy

$$\mathcal{V}_{t+1}(d^{1:t}) = \min_{(f(a_{t'}|d^{1:t'-1}))_{t'=t}^T} E\left[\sum_{\tau=t+1}^T x_{\tau} \middle| d^{1:t}\right].$$

Nyní hledáme optimální rozhodovací strategii

$$O(f(a_t|d^{1:t-1}))_{t=1}^T \in \operatorname{argmin}_{(f(a_t|d^{1:t-1}))_{t=1}^T} E\left[\sum_{t=1}^T x_t\right].$$

Protože platí

$$E\left[\sum_{t=1}^T x_t\right] = E\left[E\left[\sum_{t=1}^T x_t \middle| d^{1:T-1}\right]\right] = E\left[\sum_{t=1}^{T-1} x_t + E[x_T | d^{1:T-1}]\right],$$

můžeme psát, že

$$\begin{aligned} \min_{(f(a_t|d^{1:t-1}))_{t=1}^T} E \left[\sum_{t=1}^T x_t \right] &= \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-1}} \min_{f(a_T|d^{1:T-1})} E \left[\sum_{t=1}^T x_t \right] = \\ &= \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-1}} \min_{f(a_T|d^{1:T-1})} E \left[\sum_{t=1}^{T-1} x_t + E[x_T|d^{1:T-1}] \right]. \end{aligned} \quad (20)$$

Vnitřní střední hodnotu počítáme vzhledem k $f(d_T|d^{1:T-1})$, vnější pak vzhledem k $f(d^{1:T-1})$. Vnější střední hodnota tedy nezávisí na znáhodněné volbě a_T , proto můžeme rovnost (20) dále upravit

$$\min_{(f(a_t|d^{1:t-1}))_{t=1}^T} E \left[\sum_{t=1}^T x_t \right] = \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-1}} E \left[\sum_{t=1}^{T-1} x_t + \min_{f(a_T|d^{1:T-1})} E[x_T|d^{1:T-1}] \right]. \quad (21)$$

Označme nyní

$$\mathcal{V}_T(d^{1:T-1}) = \min_{f(a_T|d^{1:T-1})} E[x_T|d^{1:T-1}],$$

pak

$$Of(a_T|d^{1:T-1}) \in \operatorname{argmin}_{f(a_T|d^{1:T-1})} E[x_T|d^{1:T-1}].$$

Dosaďme nyní $\mathcal{V}_T(d^{1:T-1})$ do (21) a pokračujme v úpravách

$$\begin{aligned} &\min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-1}} E \left[\sum_{t=1}^{T-1} x_t + \mathcal{V}_T(d^{1:T-1}) \right] = \\ &= \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-2}} \min_{f(a_{T-1}|d^{1:T-2})} E \left[E \left[\sum_{t=1}^{T-1} x_t + \mathcal{V}_T(d^{1:T-1}) \middle| d^{1:T-2} \right] \right] = \\ &= \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-2}} \min_{f(a_{T-1}|d^{1:T-2})} E \left[\sum_{t=1}^{T-2} x_t + E[x_{T-1} + \mathcal{V}_T(d^{1:T-1})|d^{1:T-2}] \right] = \\ &= \min_{(f(a_t|d^{1:t-1}))_{t=1}^{T-2}} E \left[\sum_{t=1}^{T-2} x_t + \min_{f(a_{T-1}|d^{1:T-2})} E[x_{T-1} + \mathcal{V}_T(d^{1:T-1})|d^{1:T-2}] \right] \end{aligned}$$

Nyní označme

$$\mathcal{V}_{T-1}(d^{1:T-2}) = \min_{f(a_{T-1}|d^{1:T-2})} E[x_{T-1} + \mathcal{V}_T(d^{1:T-1})|d^{1:T-2}],$$

pak

$$Of(a_{T-1}|d^{1:T-1}) \in \underset{f(a_{T-1}|d^{1:T-2})}{\operatorname{argmin}} E[x_{T-1} + \mathcal{V}_T(d^{1:T-1})|d^{1:T-2}].$$

Takto bychom pokračovali až po $t = 1$ a získali tak rekurzivní vztah pro výpočet Bellmanových funkcí. Definujeme-li pro $t = T + 1$

$$\mathcal{V}_{T+1}(d^{1:T}) = 0, \quad (22)$$

můžeme rekurzivní vztah zapsat jako

$$\mathcal{V}_t(d^{1:t-1}) = \min_{f(a_t|d^{1:t-1})} E[x_t + \mathcal{V}_{t+1}(d^{1:t})|d^{1:t-1}] \quad (23)$$

a

$$Of(a_t|d^{1:t-1}) \in \underset{f(a_t|d^{1:t-1})}{\operatorname{argmin}} E[x_t + \mathcal{V}_{t+1}(d^{1:t})|d^{1:t-1}].$$

Bellmanovy funkce a zároveň jednotlivé podmíněné hustoty tvořící rozhodovací strategii tedy počítáme zpětným způsobem, tzn. od konce časového horizontu po jeho začátek.

Střední hodnotu v (23) počítáme vzhledem k

$$f(d_t|d^{1:t-1}) = f(x_t|a_t, d^{1:t-1})f(a_t|d^{1:t-1}) = f(a_t|d^{1:t-1}) \int f(x_t|a_t, d^{1:t-1}, \theta) f(\theta|d^{1:t-1}) d\theta,$$

kde $f(a_t|d^{1:t-1})$ je rozhodovací pravidlo, které hledáme, a $f(x_t|a_t, d^{1:t-1})$ vyjádříme pomocí modelu a aposteriorní hustoty (19), přičemž parametry θ_{jk} jsou vzhledem k aposteriorní hustotě nezávislé:

$$\begin{aligned} f(x_t = 1|a_t = k, x_{t-1} = j, a_{t-1}, d^{1:t-2}) &= \int \theta_{jk} f(\theta_{jk}|d^{1:t-1}) d\theta_{jk} = \\ &= \int \theta_{jk} \frac{\theta_{jk}^{\alpha_{jk} + A_{jk}(t) - 1} (1 - \theta_{jk})^{\beta_{jk} + B_{jk}(t) - 1}}{B(\alpha_{jk} + A_{jk}(t), \beta_{jk} + B_{jk}(t))} d\theta_{jk} = \\ &= \frac{\Gamma(\alpha_{jk} + A_{jk}(t) + 1) \Gamma(\beta_{jk} + B_{jk}(t))}{\Gamma(\alpha_{jk} + A_{jk}(t) + 1 + \beta_{jk} + B_{jk}(t))} \frac{\Gamma(\alpha_{jk} + A_{jk}(t) + \beta_{jk} + B_{jk}(t))}{\Gamma(\alpha_{jk} + A_{jk}(t)) \Gamma(\beta_{jk} + B_{jk}(t))} = \\ &= \frac{(\alpha_{jk} + A_{jk}(t)) \Gamma(\alpha_{jk} + A_{jk}(t)) \Gamma(\alpha_{jk} + A_{jk}(t) + \beta_{jk} + B_{jk}(t))}{(\alpha_{jk} + A_{jk}(t) + \beta_{jk} + B_{jk}(t)) \Gamma(\alpha_{jk} + A_{jk}(t) + \beta_{jk} + B_{jk}(t)) \Gamma(\alpha_{jk} + A_{jk}(t))} = \\ &= \frac{\alpha_{jk} + A_{jk}(t)}{\alpha_{jk} + A_{jk}(t) + \beta_{jk} + B_{jk}(t)}. \end{aligned}$$

5.4 Konkrétní příklad

5.4.1 Zadání

Uvažujme model z kapitoly 5.1. Necht' jsou skutečné pravděpodobnosti θ pro hod jedničky touto mincí rovny

$$\theta = \begin{bmatrix} \theta_{00} = 0,3 \\ \theta_{01} = 0,5 \\ \theta_{10} = 0,2 \\ \theta_{11} = 0,25 \end{bmatrix}.$$

Předpokládejme, že máme o tomto systému apriorní informaci takovou, že pravděpodobnosti hodu 1, tedy θ_{jk} , reprezentujeme z naší znalosti tak, že mají beta rozdělení s parametry α_{jk}, β_{jk} , kde

$$\begin{bmatrix} \alpha_{00} = 3 & \beta_{00} = 3 \\ \alpha_{01} = 3 & \beta_{01} = 4 \\ \alpha_{10} = 2 & \beta_{10} = 5 \\ \alpha_{11} = 1 & \beta_{11} = 2 \end{bmatrix}.$$

Tato apriorní informace odpovídá přibližně 19 předchozím pozorováním hodů touto mincí, uvažujeme-li jako výchozí apriorní hustotu Jeffreysovu apriorní hustotu (viz 4.2.3).

Nalezneme pro 8 kroků dopředu rozhodovací strategii

$$(f(a_t | d^{1:t-1}))_{t=1}^8$$

takovou, aby očekávaná ztráta

$$E \left(\sum_{t=1}^8 x_t \right)$$

byla co nejmenší.

5.4.2 Algoritmus

Pro implementaci algoritmu hledání optimální rozhodovací strategie jsme použili skriptovací programovací jazyk *Python*¹ rozšířený o balíček *NumPy*² pro operace s maticemi.

Nyní si popíšeme části algoritmu.

Nejprve programu zadáme vstupy, jedná se o

- délku rozhodovacího horizontu T ,

¹Dostupný z www.python.org.

²Dostupný z www.numpy.org.

- parametry apriorního rozdělení, reprezentovanou maticí $apri$, kterou zadáváme ve tvaru:

$$apri = np.matrix(' \alpha_{00}, \beta_{00}; \alpha_{01}, \beta_{01}; \alpha_{10}, \beta_{10}; \alpha_{11}, \beta_{11} ');$$

- a počáteční paměť x_0 pro čas $t = 1$ (výchozí stav, výsledek hodu mince před prvním hodem rozhodovacího problému).

Na jejich základě si program vypočte počet možných trajektorií a vytvoří počet v , který naplní vektory, jeden pro každý časový okamžik, sloužícími pro ukládání hodnot Bellmanových funkcí pro jednotlivé trajektorie.

```
T = 8; # rozhodovací horizont
#apri = np.matrix('1,2;2,4.01;3,3;3,3') ;
apri = np.matrix(' 3,3;3,4;2,5;1,2 '); # apriorni rozdeleni
x0 = 0; # pocatecni vstup, xprev pro t=1
v = []; # list pro ukladani Bellmanovych funkcí

for i in range(T+1):
    TT = 4*(i+1);
    v.append(np.zeros(TT));
```

Výpis 1: Vstupní parametry algoritmu pro hledání optimální strategie

Velikost vektoru v pro čas t je rovna počtu všech možných trajektorií, kterými může systém při svém chodu projít. Každý index tedy reprezentuje jednu konkrétní trajektorii. Vektor v pro čas $T + 1$ je vytvořen jako vektor samých 0, což je počáteční hodnota Bellmanovy funkce pro danou trajektorii v čase $T + 1$, viz (22). Následující funkce nám slouží k tomu, abychom pro daný index a čas t získali trajektorii ve tvaru $(a_1, x_1, \dots, a_t, x_t)'$.

```
def get_trajectory(index,t):
    """
    index - index trajektorie
    t - cas
    vystupem trajektorie jako retezec a1,x1,a2,x2,..., at, xt
    """
    n = t*2;
    out = str(bin(index)) [2:]. zfill (n);
    return out;
```

Výpis 2: Funkce pro převod indexu vektoru v pro čas t na trajektorii

Následující funkce reprezentuje fázi učení, slouží k výpočtu parametrů aposteriorního rozdělení pro danou trajektorii a čas.

```
def get_aposteriorno(index,t,x0,apriorno):
    """
    index - index trajektorie
    t - cas
    x0 - pocatecni vstup
    apriorno - parametry apriornihorozdeleni
    vystupem parametry aposteriornihorozdeleni v case t
```

```

"""
    """
    t0 = t-1;
    traj = get_trajectory (index,t0);
    if t0 == 0:
        apo = np.copy(aprioro);
    else:
        apo = np.copy(aprioro);
        for i in range(t0):
            if i == 0:
                xprev = x0;
                at = int( traj [0] );
                xt = int( traj [1] );
                if xprev == 0:
                    apo[at][0] += xt;
                    apo[at][1] += (1-xt);
                else:
                    apo[2+at][0] += xt;
                    apo[2+at][1] += (1-xt);
            else:
                ii = 2*i;
                xprev = int( traj [ ii - 1] );
                at = int( traj [ ii ] );
                xt = int( traj [ ii + 1] );
                if xprev == 0:
                    apo[at][0] += xt;
                    apo[at][1] += (1-xt);
                else:
                    apo[2+at][0] += xt;
                    apo[2+at][1] += (1-xt);
        return apo;

```

Výpis 3: Funce učení počítající parametry aposteriorního rozdělení

A zde již samotná fáze návrhu optimální strategie spočívající ve výpočtu Bellmanových funkcí.

```

for i in range(T):
    t = T-i;
    j = (4*t)/2;
    for y in range(j):
        index = y*2;
        traj = get_trajectory (index,t);
        apo = np.copy(get_aposteriornost(index,t,x0,apri));
        if t == 1:
            xprev = x0;
        else:
            xprev = int( traj [( t*2)-1] );
        at = int( traj [( t*2)-2] );
        a = v[t ][( index*4)];
        b = v[t ][( index*4)+2];
        vmin1 = min(a,b);
        c = v[t ][( index*4)+4];
        d = v[t ][( index*4)+6];
        vmin2 = min(c,d);

```

```

if xprev == 0:
    loss = float(float(apo[at][0]) / (float(apo[at][0]) + float(apo[at][1])));
    v[t-1][index] += (loss + ((1-loss)*vmin1) + (loss*vmin2));
else:
    loss = float(float(apo[2+at][0]) / (float(apo[2+at][0]) + float(apo[2+at][1])));
    v[t-1][index] += (loss + ((1-loss)*vmin1) + (loss*vmin2));

```

Výpis 4: Algoritmus hledání optimální strategie

5.4.3 Optimální strategie

Pro daná vstupní data navrhl program za optimální strategii posloupnost akcí

$$a_1 = 0, a_2 = 1, \text{ pro } x_1 = 0, a_2 = 0, \text{ pro } x_1 = 1, \dots,$$

atd.

Simulujme nyní výsledky hodů touto mincí, přičemž zvolíme posloupnost rozhodovacích pravidel navrženou optimální strategií. Hody budeme simulovat celkem 1000krát.

Samotnou simulaci výsledku hodu mince jsme provedli pomocí následujícího tvrzení. Necht' θ značí pravděpodobnost hodu jedničky, tedy

$$P(x = 1) = \theta.$$

Necht' u je náhodná veličina z rovnoměrného rozdělení na intervalu $(0, 1)$

$$u \sim R(0, 1)$$

Pak náhodný jev

$$u < \theta$$

přísluší výsledku hodu mince $x = 1$ a

$$u \geq \theta$$

naopak $x = 0$. Pak

$$P(x = 1) = P(u < \theta) = \theta$$

a tedy můžeme tyto náhodné jevy považovat za totožné.

Simulace je realizována pomocí následujícího kódu.

```

p = 1000; # pocet realizaci
a = np.copy(apri);
theta = [0.3, 0.5, 0.2, 0.25]; # skutečne pravdepodobnosti hodu 1

realizace = np.zeros(shape=(p,T)); # vektor pro výsledky realizace p v case t
ztrata = np.zeros(p); # vektor pro ukládání ztráty realizace p

def min2(a,b):

```

```

#####
a=hodnota_Bellmanovy_funkce_pro_akci_0
b=hodnota_Bellmanovy_funkce_pro_akci_1
vystup=akce_0/1_dle_mensi_z_hodnot
#####

min = 0;
if b < a:
    min = 1;
return min

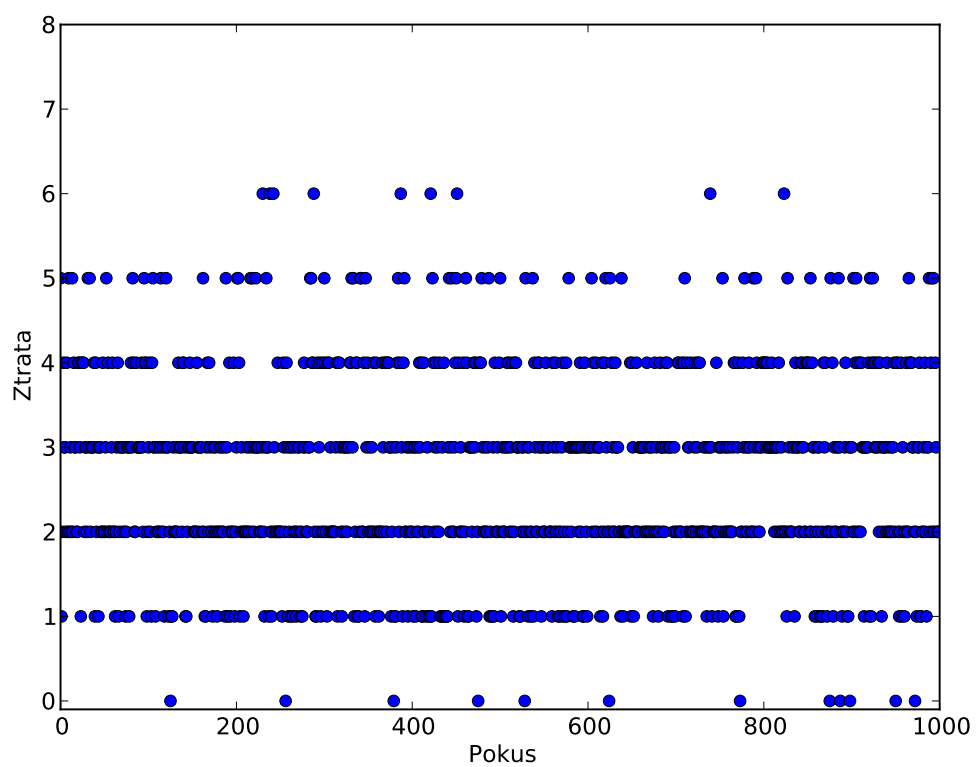
# samotna realizace
for i in range(p):
    index = 0;
    for j in range(T):
        if j == 0:
            xprev = x0;
            at = min2(v[0][0], v[0][2]) ;
        else:
            xprev = realizace[i][j-1];
            at = min2(v[j][index], v[j][index+2])
        if xprev == 0:
            pst = theta[at];
        else:
            pst = theta[2+at];
        u = np.random.uniform(0,1);
        if u < pst:
            realizace[i][j] = 1;
        else:
            realizace[i][j] = 0;
        if at == 0:
            if realizace[i][j] == 0:
                index += 0;
            else:
                index += 1;
        else:
            if realizace[i][j] == 0:
                index += 2;
            else:
                index += 3;
        index = index*4;
        ztrata[i] = np.sum(realizace[i]);

# vypocet prumerne ztraty
celkova = np.sum(ztrata);
print 'Ztrata: ';
print celkova;
prumer = celkova / p;
print prumer;

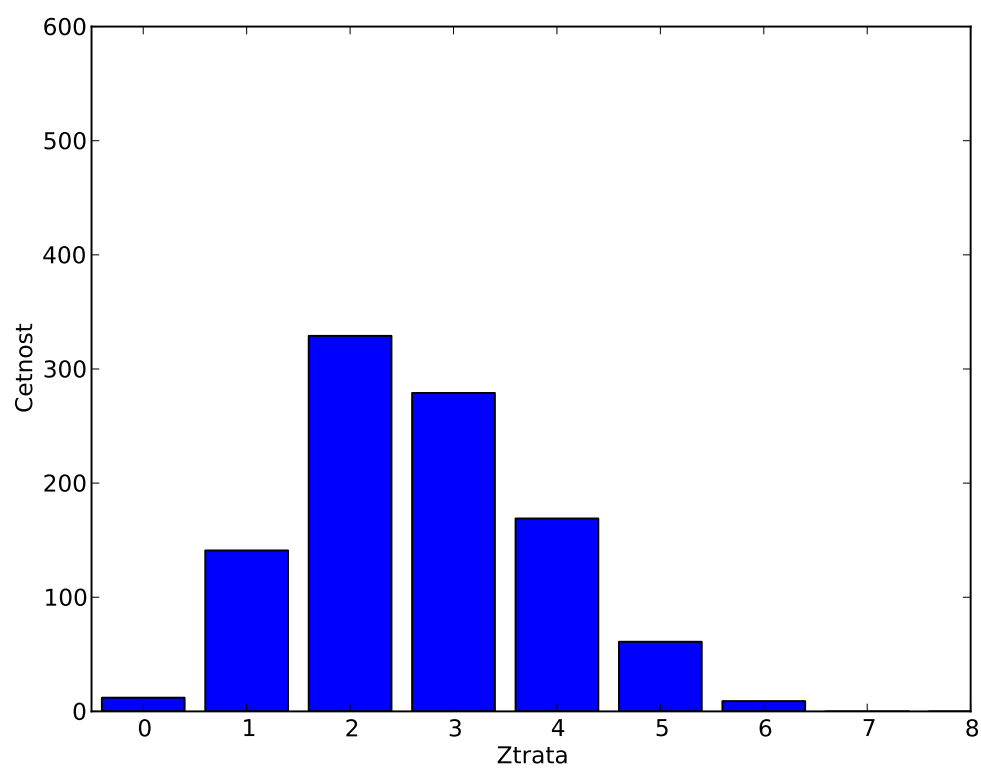
```

Výpis 5: Simulování výsledků hodů mincí

Při realizaci a dodržování navržené optimální strategie došlo k průměrné ztrátě 2, 671. Na obrázku 6 jsou zobrazeny ztráty jednotlivých pozorování, na obrázku 7 pak četnosti jednotlivých ztrát při 1000 realizacích.



Obrázek 6: Ztráty optimální strategie v jednotlivých pokusech pro 1000 simulací.



Obrázek 7: Četnosti ztrát v 1000 simulovaných pokusech optimální strategie.

5.4.4 Porovnání se suboptimální strategií

Vzhledem k tomu, že hledání optimální strategie může i pro jednoduchý systém vést k výpočetně náročným úlohám, aproximujeme danou rozhodovací úlohu hledáním suboptimálních strategií.

Uvažujme suboptimální strategii, která se rozhoduje taktéž s ohledem na všechny trajektorie až do konce časového horizontu, ovšem při návrhu strategie nevykonává fázi učení, tedy při řízení nebere v potaz informace, které o systému získá.

Porovnejme optimální strategii s touto suboptimální, která při svém řízení nebere v úvahu fázi učení se o systému. Rozdíl v algoritmu suboptimální strategie oproti algoritmu optimální strategie 5.4.2 je ten, že očekávaný přírůstek ztráty v daném čase pro danou trajektorii počítá místo vzhledem k aposteriornímu rozdělení naopak vzhledem k apriornímu rozdělení.

Při použití této suboptimální strategie je pak navržena posloupnost akcí

$$a_1 = 1, a_2 = 1, \text{ pro } x_1 = 0, a_2 = 1, \text{ pro } x_1 = 1, \dots,$$

atd.

Při simulaci hodů mincí s ohledem na volbu akcí navržených suboptimální strategií by naše průměrná ztráta byla o něco vyšší, a to 3,272. Ztráty jednotlivých pokusů při volbě této strategie jsou zobrazeny na obrázku 8, na obrázku 9 jsou pak četnosti jednotlivých ztrát při 1000 realizací. Srovnání četností ztrát pro optimální a suboptimální strategie jsou pak na obrázku 10. Z posledního obrázku je patrné, že optimální strategie je o poznání lepší.

5.5 Optimální strategie navrhující krátkodobě nevýhodná rozhodnutí

Vzhledem k povaze duálního řízení se může stát, že řídicí systém v určitých časech navrhne rozhodnutí, které je z krátkodobého hlediska nevýhodné. To znamená, že pravděpodobnost hodu panny je pro navrženou akci v daném okamžiku nižší, než pro akci jinou. Zato však o systému získá cennou informaci, která může zlepšit řízení v následujících krocích a celkově vést k lepšímu zisku, resp. k menším ztrátám.

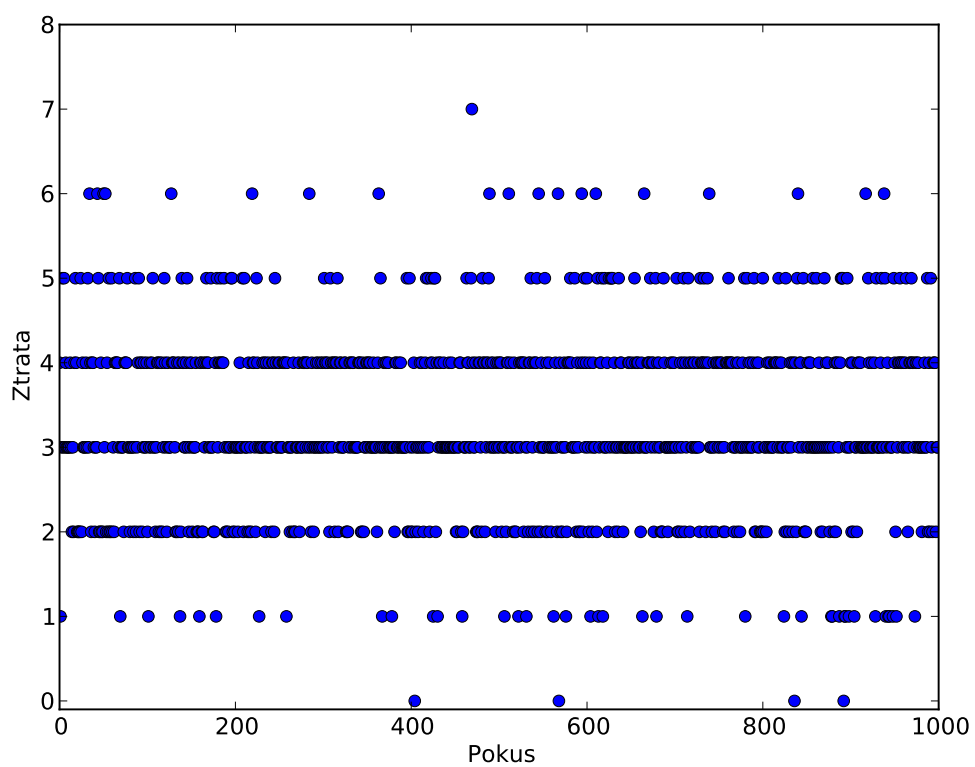
Cílem této kapitoly je nalézt takovou apriorní distribuci, která povede k tomu, že optimální strategie navrhne právě takováto krátkodobě nevýhodná rozhodnutí.

Poznamenejme však, že následující postup slouží pouze k ilustrování výše popsané situace, v praktických úlohách nemá volba apriorního rozdělení následujícím způsobem uplatnění.

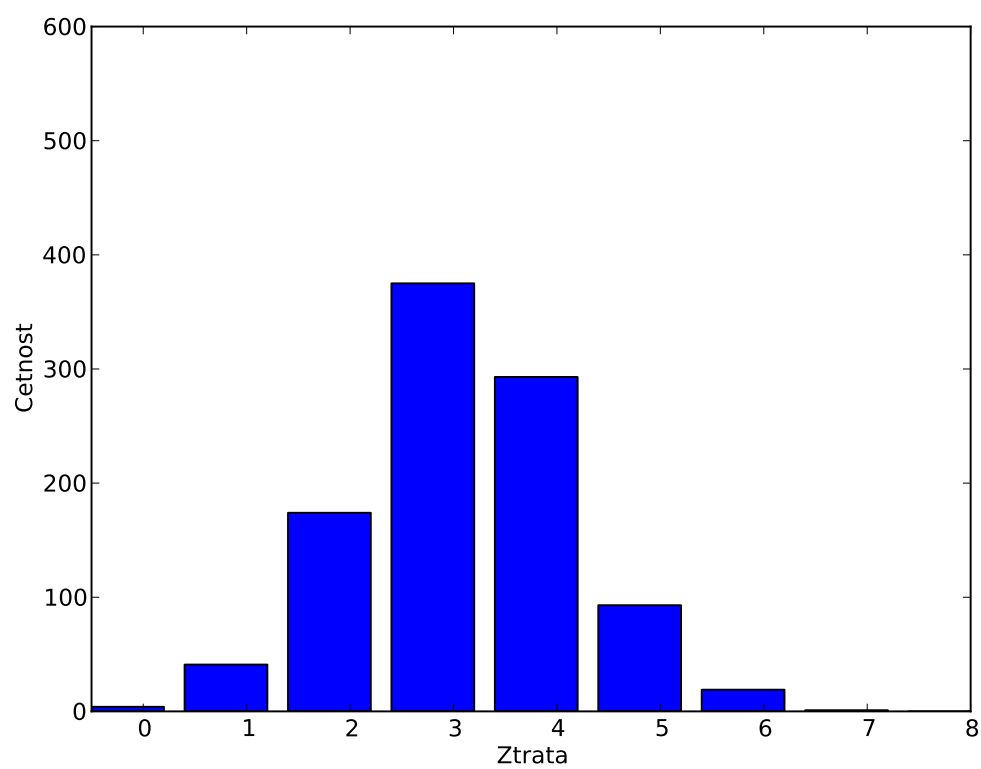
Nalezněme takovou apriorní distribuci pro řídicí systém navrhující optimální strategii na dva kroky dopředu, tedy $T = 2$, tak, aby se optimální strategie rozhodla v prvním kroku jinak, než strategie, která neuvažuje fázi učení.

Nechť

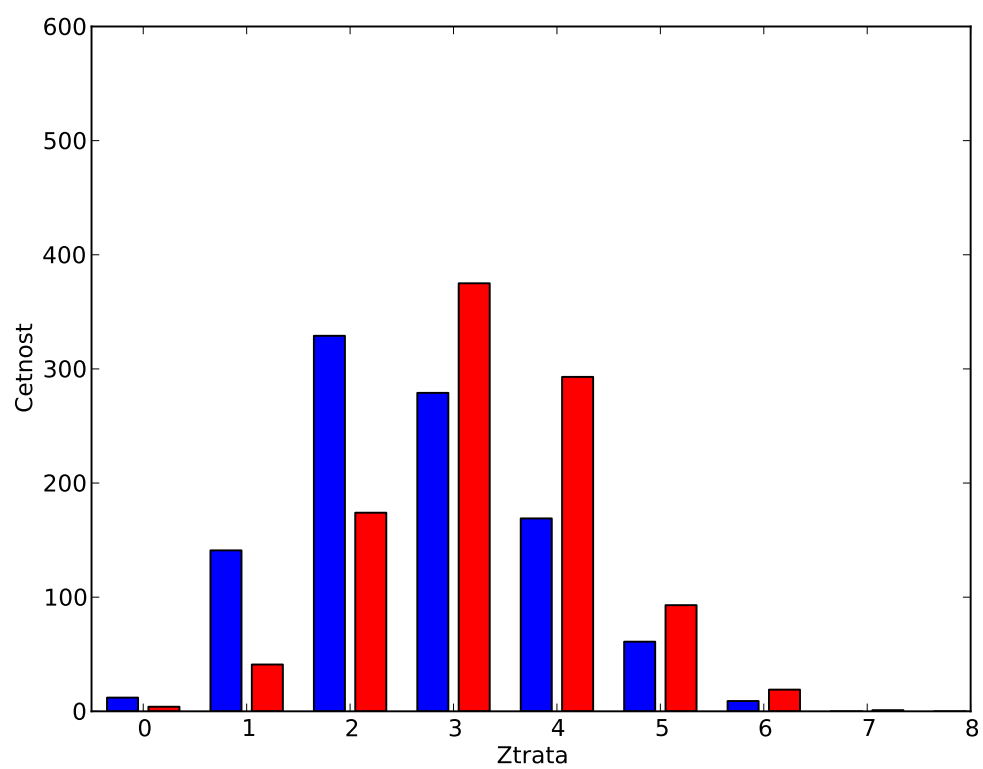
- $\theta = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})'$ je pravděpodobnost hodu 1,
- $\theta_{jk} \sim Be(\alpha_{jk}, \beta_{jk})$, kde $\forall j, k : \alpha_{jk}, \beta_{jk} \in \mathbb{R}^+$



Obrázek 8: Ztráty suboptimální strategie v jednotlivých pokusech pro 1000 simulací.



Obrázek 9: Četnosti ztrát v 1000 simulovaných pokusech suboptimální strategie.



Obrázek 10: Srovnání četností ztrát v 1000 simulovaných pokusech optimální a suboptimální strategie.

- $f(\theta) \propto \prod_{ij} \theta_{ij}^{\alpha_{ij}-1} (1 - \theta_{ij})^{\beta_{ij}-1}$
- a $x_0 = 0$.

Pro jednoduchost předpokládejme, že

$$\alpha_{10} = \beta_{10} = \alpha_{11} = \beta_{11}.$$

Nyní chceme, aby při volbě akce 0 v čase $t = 1$ byla pravděpodobnost ztráty, pokud neuvažujeme učení, větší, než při volbě akce 1, tedy

$$\begin{aligned} & \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} + \left(1 - \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}\right) \min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \frac{1}{2} > \\ & > \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} + \left(1 - \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) \min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) + \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \frac{1}{2}. \end{aligned}$$

Jednoduchými úpravami

$$\begin{aligned} & \frac{3}{2} \left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} - \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \right) > \min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) \left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} - \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \right) \\ & \left(\frac{3}{2} - \min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) \right) \left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} - \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \right) > 0, \end{aligned}$$

a protože $\forall \alpha_{00}, \beta_{00}, \alpha_{01}, \beta_{01} \in \mathbb{R}^+$ platí

$$\frac{3}{2} - \min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) > 0,$$

dospějeme k závěru, že musí platit podmínka

$$\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} > \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}. \quad (24)$$

Zvolme α_{00}, β_{00} libovolně. Abychom splnili podmínku (24), volíme

$$\begin{aligned} \alpha_{01} &= k\alpha_{00}, \\ \beta_{01} &= k\beta_{00} + \varepsilon, \end{aligned}$$

kde $k \in \mathbb{R}^+, \varepsilon \in \mathbb{R}^+$. Omezme se na $\varepsilon \in (0, 1)$, aby platilo, že

$$\min\left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + 1}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}\right) = \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + 1}. \quad (25)$$

Z předpokladu (24) zřejmě plyne

$$\min \left(\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}}, \frac{\alpha_{01}}{\alpha_{01} + \beta_{01} + 1} \right) = \frac{\alpha_{01}}{\alpha_{01} + \beta_{01} + 1}. \quad (26)$$

Nyní chceme, aby celková očekávaná ztráta optimální strategie na 2 kroky dopředu byla při volbě akce 0 v prvním kroku nižší, tedy (využijeme-li (25), (26))

$$\begin{aligned} & \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} + \left(1 - \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \right) \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + 1} + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \frac{1}{2} < \\ & < \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} + \left(1 - \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \right) \frac{\alpha_{01}}{\alpha_{01} + \beta_{01} + 1} + \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}} \frac{1}{2}. \end{aligned}$$

Dosadíme za α_{01} a β_{01} a dostáváme

$$\begin{aligned} & \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} + \left(1 - \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \right) \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + 1} + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \frac{1}{2} < \\ & < \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + \frac{\varepsilon}{k}} + \left(1 - \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + \frac{\varepsilon}{k}} \right) \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + \frac{\varepsilon}{k} + \frac{1}{k}} + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00} + \frac{\varepsilon}{k}} \frac{1}{2}. \end{aligned}$$

Tedy pro $\varepsilon \rightarrow 0$ můžeme volit $k > 1$.

5.5.1 Ukázka na konkrétním příkladu

Do programu zadejme následující parametry apriorního rozdělení (beta) pro pravděpodobnosti hodu jedničky

$$apri = \begin{bmatrix} \alpha_{00} = 1 & \beta_{00} = 2 \\ \alpha_{01} = 2 & \beta_{01} = 4,01 \\ \alpha_{10} = 3 & \beta_{10} = 3 \\ \alpha_{11} = 3 & \beta_{11} = 3 \end{bmatrix},$$

které splňují podmínky určené dříve v této kapitole.

Program určí jako optimální posloupnost akcí

$$a_1 = 0, a_2 = 0, \text{ pro } x_1 = 0, a_2 = 0, \text{ pro } x_1 = 1.$$

V prvním kroku se zřejmě rozhodl z krátkodobého hlediska nevýhodně, protože pravděpodobnost hodu 1 je při zvolené akci 0 vyšší:

$$\frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} = \frac{1}{3} > \frac{2}{6,01} = \frac{\alpha_{01}}{\alpha_{01} + \beta_{01}}.$$

Rozhodl se vzhledem k tomu, že se o systému již po první realizaci naučil novou informaci, která ho s ohledem na budoucí vývoj donutí toto krátkodobě nevýhodné rozhodnutí provést.

Skutečné p-stí hodů 1 ($\theta_{00}; \theta_{01}; \theta_{10}; \theta_{11}$)	Parametry apriorního rozdělení ($\alpha_{00}, \beta_{00}; \alpha_{01}, \beta_{01}; \alpha_{10}, \beta_{10}; \alpha_{11}, \beta_{11}$)	PZO	PZS
(0, 4; 0, 6; 0, 6; 0, 3)	(4, 6; 6, 4; 6, 4; 3, 7)	3, 882	3, 919
(0, 4; 0, 6; 0, 6; 0, 3)	(13, 17; 17, 10; 7, 3; 10, 25)	3, 791	3, 800
(0, 4; 0, 6; 0, 6; 0, 3)	(51, 74; 60, 42; 55, 36; 47, 70)	3, 863	3, 866
(0, 4; 0, 6; 0, 6; 0, 3)	(45, 57; 60, 43; 50, 35; 35, 75)	3, 901	3, 868
(0, 7; 0, 3; 0, 4; 0, 8)	(6, 3; 4, 8; 6, 5; 3, 1)	4, 186	4, 229
(0, 7; 0, 3; 0, 4; 0, 8)	(2, 1; 1, 4; 2, 2; 3, 1)	4, 153	4, 267
(0, 7; 0, 3; 0, 4; 0, 8)	(20, 10; 10, 40; 20, 20; 30, 10)	4, 208	4, 234
(0, 6; 0, 7; 0, 5; 0, 4)	(4, 6; 4, 2; 5, 2; 2, 7)	4, 420	4, 485

Tabulka 3: Průměrné ztráty optimální strategie (PZO) a suboptimální strategie (PZS) pro různé skutečné parametry mince a zvolené parametry apriorního beta rozdělení.

Suboptimální strategie, kterou jsme uvažovali i v předchozí podkapitole, tedy neučící se strategie, navrhne naopak strategii

$$a_1 = 1, a_2 = 1, \text{ pro } x_1 = 0, a_2 = 1, \text{ pro } x_1 = 1.$$

5.6 Další příklady

V tabulce 3 jsou zobrazeny průměrné ztráty optimální a suboptimální strategie pro různé skutečné parametry mince a různě zvolené parametry apriorního beta rozdělení.

Strategie jsou konstruovány pro $T = 8$ a průměrná ztráta vypočtena z 1000 simulovaných realizací.

Z dat obsažených v této tabulce můžeme vyčíst, že ve většině případů vede optimální strategie k alespoň mírně menší průměrné ztrátě. Při přesnějším apriorním rozdělení (reprezentujícího např. více apriorních testovacích pokusů hodů mincí) se pak rozdíl v průměrné ztrátě mezi optimální a suboptimální strategií zmenšuje. V jednom případě se suboptimální strategie projevila jako stejně dobrá a dokonce i mírně lepší. Důvodem může být nepřesnost apriorní informace, minimální rozdíl pak může záviset i na náhodě při konkrétních simulacích.

6 Závěr

V průběhu této práce jsme se seznámili s problematikou rozhodování za neurčitosti. Získané znalosti jsme poté uplatnili při sestavení algoritmu hledajícího optimální rozhodovací strategii jednoduchého Markovského rozhodovacího procesu závislého na kombinaci vstupu a paměti. Tento proces jsme reprezentovali opakovanými hody nesymetrickou mincí, kde neurčitost byla dána naší nepřesnou znalostí parametrů této mince, tedy pravděpodobností výsledků hodů touto mincí pro jednotlivé kombinace vstupu a paměti.

Optimální strategii jsme poté na pár konkrétních příkladech porovnali s konkrétní suboptimální strategií. A to tak, že jsme simulovali hody touto mincí při dodržování navržené optimální a suboptimální strategie. Na výsledcích těchto simulací lze pozorovat, že je optimální strategie skutečně lepší než zvolená suboptimální strategie.

Vzhledem k tomu, že jsou při hledání optimální strategie uvažovány všechny možné trajektorie, kterými může systém projít, čelíme problému s výpočetní složitostí algoritmu a jeho nároky na paměť. I pro náš relativně jednoduchý model mince pak může systém v konkrétním čase t projít celkem 4^t různými trajektoriemi. Pro každou trajektorii je pak v každém čase počítána Bellmanova funkce.

Optimalizace algoritmu tím způsobem, že bychom zohlednili třídy ekvivalencí jednotlivých trajektorií, by vedla ke snížení nároků na potřebnou paměť. Jednotlivé třídy ekvivalence by pak tvořily takovou množinu trajektorií, které obsahují stejný počet konkrétních přeskoků řetězce a které končí stejným výstupem, resp. vstupem pro další časový okamžik.

7 Reference

- [1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 1985.
- [2] J. M. Bernardo a A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 2. vydání, 1997.
- [3] R. Briš a M. Litschmannová. *Statistika II*. [online], Ostrava 2007. Dostupné z: <http://homel.vsb.cz/~bri10/Teaching/Statistika%20II/skriptum/>
- [4] Marie Hušková. *Bayesovské metody*. Univerzita Karlova, SPN Praha, 1. vydání, 1985.
- [5] J. Kracík. *Cooperation Methods in Bayesian Decision Making with Multiple Participants*. [online] Dostupné z: <http://tinyurl.com/kracik-phd-thesis>
- [6] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma a L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer, London, 2005.
- [7] M. Litschmannová. *Úvod do statistiky*. [online] , Ostrava 2011. Dostupné z: <http://mi21.vsb.cz/modul/uvod-do-statistiky>
- [8] M. Litschamnnová. *Vybrané kapitoly z pravděpodobnosti*. [online], Ostrava 2011. Dostupné z: <http://mi21.vsb.cz/modul/vybrane-kapitoly-z-pravdepodobnosti>
- [9] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.
- [10] A. A. Feldbaum. *Optimal control systems*. Academic Press, New York, 1965.